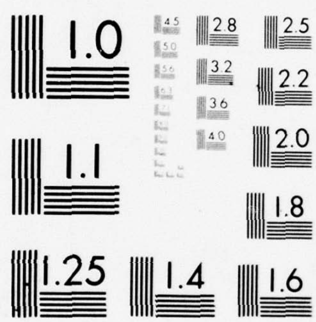
AD-A042 661 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
ON DIFFERENTIALS, ASYMPTOTIC NORMALITY AND ALMOST SURE BEHAVIOR--ETC(U)
MAY 77 D D BOOS, R J SERFLING N00014-76-C-0608
UNCLASSIFIED FSU-STATISTICS-M415 ONR-TR-115 NL

1 OF 1

ADA042 661

END
DATE
FILMED
8-77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

e Florida State University

(12)

(6)

ON DIFFERENTIALS, ASYMPTOTIC NORMALITY AND ALMOST SURE BEHAVIOR OF STATISTICAL FUNCTIONS, WITH APPLICATION TO M-STATISTICS FOR LOCATION PARAMETERS

by Dennis D. Boos [redacted] R. J. Serfling

(10) FSU-Statistics [redacted] M415

(14) ONR Technical Report No. 115

(18) ONR

(19) TR-115

(9) Technical rept.,

(11) May [redacted]77

(12) 79p.

Department of Statistics
Florida State University
Tallahassee, Florida 32306

(15)

Research supported by the Army, Navy and Air Force under Office of Naval Research Contract No. N00014-76-C-0608. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ABSTRACT

ON DIFFERENTIALS, ASYMPTOTIC NORMALITY AND ALMOST SURE BEHAVIOR OF STATISTICAL FUNCTIONS, WITH APPLICATION TO M-STATISTICS FOR LOCATION PARAMETERS

Parameters of interest in statistics can often be expressed as functionals $T(F)$ of the underlying population distribution function, in which case a natural sample analogue estimator is provided by the "statistical function" $T(F_n)$ based upon the sample distribution function F_n.

Several notions of differentiability of functionals T are formulated, including innovations designed to broaden the scope of statistical application. Methodology for finding the differential, and for utilizing it to characterize the asymptotic distribution and almost sure behavior of statistical functions, is presented. Typically this means asymptotic normality and the law of the iterated logarithm. Previous work of von Mises (1947), Kallianpur and Rao (1955), Filippova (1962), Gregory (1976) and Beran (1977) is relevant.

Application to M-estimates for location parameters is carried out. The solution of the equation $\int\psi(x - T(F))dF(x) = 0$ is formulated as an M-functional and conditions for its differentiability are investigated. Asymptotic normality and the law of the iterated logarithm for $T(F_n)$ are established under regularity conditions on ψ slightly stronger than continuity and under minimal restrictions on F. One-step estimators and the case of scale unknown are also treated. Previous work of Huber (1964), Hampel (1974), Carroll (1975, 1977), Collins (1976), Portnoy (1977) and Beran (1977) is augmented.

CONTENTS

1. Introduction. Consider a functional $T(\cdot)$ defined on distribution functions G, for example the variance functional $T(G) = \int[x - \int x dG(x)]^2 dG(x)$. Corresponding to a sample $X_1, \ldots, X_n$ from a distribution F, let F_n denote the sample distribution function (see (2.16)). The "statistical function" $T(F_n)$ is the natural sample analogue estimator of the "parameter" $T(F)$.

The first part of this paper treats the differentiability of functionals T and shows how it forms the basis of a methodology for obtaining the convergence theory of statistical functions $T(F_n)$. Specifically, it is seen how to characterize the asymptotic distribution and almost sure behavior of $T(F_n)$. Typically, this means *asymptotic normality* and the *law of the iterated logarithm*:

$$(1.1) \qquad \sqrt{n}\,[T(F_n) - T(F) - \mu(T, F)] \xrightarrow{d} N(0, \sigma^2(T, F)),\ n \to \infty;$$

$$(1.2) \qquad \overline{\lim_{n\to\infty}} \frac{\sqrt{n}\,[T(F_n) - T(F) - \mu(T, F)]}{\sqrt{2\sigma^2(T, F)\log\log n}} = 1 \quad \text{w.p.1.}$$

Of course, when $\mu(T, F) = 0$, (1.2) implies *strong consistency* of $T(F_n)$. Here, convergence in distribution is denoted by $\xrightarrow{d}$, "with probability 1" is denoted by "w.p.1," and $N(m, v)$ denotes a random variable having the normal distribution with mean m and variance v. The quantity $\mu(T, F)$ represents an asymptotic bias parameter (usually 0) and the quantity $\sigma^2(T, F)$ an asymptotic variance parameter.

The functional representation of statistical parameters was first studied in detail by von Mises (1947), who developed a theory of differentiation of statistical functions $T(F_n)$ and employed corresponding Taylor expansions as a tool for investigation of $T(F_n)$. This work was extended in the framework of stochastic process theory by Filippova (1962). A recasting of von Mises' approach in the context of Frechet differentiation was introduced by Kallianpur

and Rao (1955). This approach bypasses the troublesome remainder term in the Taylor expansion but introduces a new technical difficulty, the handling of a norm. In this vein, further development in the modern analysis context of Frechet differentiation in Banach spaces has been provided by Gregory (1976). All of these authors implement the approach to characterize the asymptotic distribution theory of $T(F_n)$ for selected types of functional $T(\cdot)$, but do not consider the question of the almost sure behavior of $T(F_n)$.

In the present development, attention is focused upon the *differential* of a functional T as the key concept and tool. Although essentially in the spirit of Frechet differentiation, our orientation differs slightly from the modern analysis treatments, in that we conveniently avoid the requirement that the domain of T be a linear space. Furthermore, we enrich the notion of differential in two ways designed to broaden the scope of its statistical application. One modification, called the quasi-differential, permits a slight degree of nonlinearity in the form of the differential. The other modification consists of stochastic versions of the (quasi-) differential. Indeed, the concept of *stochastic quasi-differential* is sufficiently flexible that we are able to formulate it even for a sequence of statistics $\{T_n\}$ not associated with a functional $T(\cdot)$.

In Section 2 we present these various notions of differential and provide methodology for finding the differential and establishing its validity as such. General discussion of the statistical role of the differential is provided, and some specific statistical applications are developed. Lemmas 2.5 and 2.7 characterize the role of the (stochastic quasi-) differential in reduction of the problem of convergence of $T(F_n) - T(F)$ to a similar problem for a random variable of standard form, i.e., a *sum* $\sum_1^n g(X_i)$. These results are given without restriction on the dependence of the X_i's. As particular applications of the lemmas, for the case of *independent* X_i's, Theorems 2.1

and 2.2 provide conclusions of the forms (1.1) and (1.2). Here there are no restrictions imposed on the underlying distribution F, other than what is implicitly required in order that T(F) be well-defined and that the differential of T(·) be defined. Thus the theorems are applicable in connection with a wide range of functionals T(·) and distributions F. As an illustration of the methodology, the sample variance statistic is considered. Concluding Section 2, some complementary discussion and details are provided.

For implementation of the differential approach, a norm plays an intermediate yet fundamental role. In specializing our methodology to obtain the afore-mentioned Theorems 2.1 and 2.2, we utilize the "sup norm" and specifically make heavy use of the random variable

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F(x)|, \tag{1.3}$$

i.e., the "Kolmogorov-Smirnov statistic," for which considerable probabilistic theory is available. In extending Theorems 2.1 and 2.2 to the case of *dependent* X_i's, it is necessary to deal only with sums of the form $\sum_1^n g(X_i)$ and with the statistic D_n. Thus such potential extension is straightforward and broad.

Augmenting the methodology of Section 2, we present in Section 3 an inequality which is useful in dealing with the sup norm.

The second part of this paper utilizes our differential approach to establish new results for M-estimates of location parameters. These estimators correspond to functionals T defined as solutions of equations of the form

$$\int \psi(x - T(F))dF(x) = 0. \tag{1.4}$$

The "M-estimate" of $T(F)$ is then $T(F_n)$. Under very broad assumptions on ψ and F, we obtain asymptotic normality and the law of the iterated logarithm for $T(F_n)$, augmenting previous work by Huber (1964), Hampel (1968), (1974), Carroll (1975), (1977), Collins (1976), Portnoy (1977), and Beran (1977a, b). For some choices of Ψ covered by earlier authors' theorems on asymptotic normality of $T(F_n)$, we in addition characterize the almost sure behavior and/or relax the assumptions on F. We also cover important choices of ψ excluded by existing theorems in the literature.

Section 4 presents this development, commencing with a discussion of varieties of ψ of interest: "Hubers," "Hampels," etc. In formulating a precise notion of the solution of (1.4) as a functional $T(F)$, it is necessary to take into account that the equation (1.4) may have multiple solutions for some choices of ψ. We formulate a version we call the "M-functional" and investigate conditions under which it possesses a stochastic quasi-differential. Our extended notions of differential are advantageous here in reducing the restrictions needed on ψ. Applying the differentiability thus established, we present in Theorem 4.5 conclusions of form (1.1) and (1.2) for $T(F_n)$, in the case of independent X_i's. The regularity conditions on ψ are slightly stronger than continuity, and the regularity conditions on F are minimal. Extension to cases of dependent X_i's would be straightforward. We also treat one-step versions f M-estimators for location, with conclusions of form (1.1) and (1.2) given by Theorem 4.8. A different type of extension, to the case of scale unknown, is given by Theorem 4.10 and its corollary. Concluding Section 4, we make comparisons with other results in the literature.

The discussion of the sample variance in Section 2 entails incidentally the random variable D_n in (1.3) and turns up the following seemingly open question, of possible interest to probabilists. Under what normalization by constants $\{a_n\}$ do we have

$$\text{(1.5)} \qquad \overline{\lim_{n\to\infty}} \frac{a_n}{D_n} = 1 \text{ w.p.1 } ?$$

(In the reciprocal case, i.e., for $a_n\ D_n$, the answer is, of course, well-known - see Lemma 2.8.)

2. Differentials of functionals and their statistical applications.

2.0. *Preliminary remarks.* Let T be a real-valued functional defined on a set $\mathcal{F}$ of distribution functions. To avoid trivial complications, assume that $\mathcal{F}$ contains for each x, $-\infty < x < \infty$, the distribution function δ_x degenerate at the value x. Also, assume that $\mathcal{F}$ is *convex*: for each G and H in $\mathcal{F}$, the "line segment" joining G and H, i.e., the set of distribution functions

$$(1 - \lambda)G + \lambda H = G + \lambda(H - G), \quad 0 \le \lambda \le 1,$$

belongs to $\mathcal{F}$. Denote by $\mathcal{D}(\mathcal{F})$ the linear space generated by *differences* H - G of members of $\mathcal{F}$. Note that $\mathcal{D}(\mathcal{F})$ may be represented as $\{\Delta: \Delta = c(G - F),$ for F and G in $\mathcal{F}$ and c real$\}$. We shall consider $\mathcal{D}(\mathcal{F})$ to be equipped with a norm $||\cdot||$.

In this section we first consider a basic notion of differential for functionals T defined on distribution functions, and we consider methodology for finding the form of the differential and verifying its validity. Next we briefly discuss the statistical role of the differential. Based on this discussion, we then introduce certain modifications of the differential designed to broaden the scope of statistical application. The notion of differential is formulated even for the case of a sequence of statistics *not* generated by any functional T. Following these preparations, specific statistical applications of the differential approach are presented. In particular, theorems pertaining to weak and strong consistency, asymptotic normality, and the law of the iterated logarithm for statistical functions are developed. An illustration for the sample variance statistic is carried out. Finally, complementary

discussion regarding possible norms and other aspects is provided, and related work by other authors is cited.

2.1. *Basic concept of differential.*

DEFINITION. We say that a functional T defined on $\mathcal{F}$ has a *differential* at the point $F \in \mathcal{F}$ with respect to the norm $||\cdot||$ and the set $\mathcal{G}_F \subset \mathcal{F}$ if there exists a quantity $T(F; \Delta)$, defined on $\Delta \in \mathcal{D}(\mathcal{F})$, which is linear in the argument Δ and satisfies the condition

$$(2.1) \qquad \lim_{\substack{||G-F|| \to 0 \\ G \in \mathcal{G}_F}} \frac{T(G) - T(F) - T(F; G - F)}{||G - F||} = 0. \ \square$$

($T(F; \Delta)$ is called the "differential.")

REMARKS. (i) By (2.1) is meant that for each $\varepsilon > 0$ there exists $\delta > 0$ such that $0 < ||G - F|| < \delta$, $G \in \mathcal{G}_F$, implies

$$(2.2) \qquad |T(G) - T(F) - T(F; G - F)| \le \varepsilon \, ||G - F||.$$

(ii) To establish (2.1) or (2.2), it suffices (see Apostol (1957), p.65) to verify that it holds for all *sequences* $\{G_n\}$ in $\mathcal{G}_F$ satisfying $||G_n - F|| \to 0$, $n \to \infty$.

(iii) By *linearity* of $T(F; \Delta)$ is meant that

$$(2.3) \qquad T(F; \sum_{i=1}^{k} a_i \Delta_i) = \sum_{i=1}^{k} a_i \, T(F; \Delta_i)$$

for $\Delta_1, \ldots, \Delta_k \in \mathcal{D}(\mathcal{F})$ and real $a_1, \ldots, a_k$.

(iv) In the general context of differentiation in Banach spaces, the differential $T(F; \Delta)$ would be called the *Frechet derivative* of T. In such treatments, the space $\mathcal{F}$ on which T is defined is assumed to be a normed *linear* space. We intentionally avoid this assumption here, in order to avoid defining the functional T at points F which are not distribution functions. $\square$

In typical cases, the quantity $T(F; G - F)$ may be characterized as a directional derivative. The *right-hand directional derivative* of the functional T at the point F in the direction of G is defined as

$$D_G\, T(F) = \lim_{t \to 0+} \frac{T(F + t(G - F)) - T(F)}{t}, \tag{2.4}$$

provided that this limit exists. Note that $D_G\, T(F)$ is just the usual right-hand derivative, at $t = 0$, of the function

$$Q(t) = T(F + t(G - F))$$

defined as a function of a real variable t, $0 \le t \le 1$, for fixed distributions F and G. That is, $D_G\, T(F) = Q_+'(0)$. We now show that if the set $\mathcal{G}_F$ is sufficiently rich, then the differential $T(F; G - F)$ is given by $D_G\, T(F)$.

DEFINITION. A set $\mathcal{G}$ of distributions is *weakly starshaped at* F *with respect to* $\mathcal{F}$ if for each $G \in \mathcal{F}$, the distributions $F_\lambda = (1 - \lambda)F + \lambda G$ belong to $\mathcal{G}$ for all sufficiently small $\lambda > 0$. □

LEMMA 2.1. *Suppose that* T *has a differential at* F *with respect to* $||\cdot||$ *and* $\mathcal{G}_F$. *Suppose that* $\mathcal{G}_F$ *is weakly starshaped at* F *with respect to* $\mathcal{F}$. *Then, for each* $G \in \mathcal{F}$, $D_G\, T(F)$ *exists and satisfies*

$$D_G\, T(F) = T(F; G - F). \tag{2.5}$$

PROOF. Note that $F_\lambda - F = \lambda(G - F)$, so that $||F_\lambda - F|| = \lambda||G - F|| \to 0$ as $\lambda \to 0+$. Utilizing (2.1), the convexity of $\mathcal{F}$, and the linearity of $T(F; \Delta)$, and taking $\lambda > 0$ sufficiently small that $F_\lambda \in \mathcal{G}_F$, we have

$$\begin{aligned} T(F_\lambda) - T(F) &= T(F; F_\lambda - F) + o(||F_\lambda - F||),\ \lambda \to 0+, \\ &= \lambda[T(F; G - F) + o(1)],\ \lambda \to 0+. \end{aligned}$$

Thus (2.5) follows. □

The significance of (2.5) is that it permits the quantity $T(F; G - F)$, which is defined only with respect to a specific norm $||\cdot||$, to be obtained as a quantity $D_G T(F)$ defined without reference to any norm. This aspect will be important in a scheme, now to be introduced, for *determination* of the differential $T(F; \Delta)$.

A special role in the present development, as well as later in the statistical application of the differential, is played by the function

$$T[F; x] = T(F; \delta_x - F) - \mu(T, F), \quad -\infty < x < \infty, \tag{2.6}$$

where

$$\mu(T, F) = \int_{-\infty}^{\infty} T(F; \delta_x - F)dF(x). \tag{2.7}$$

Note that

$$\int_{-\infty}^{\infty} T[F; x]dF(x) = 0. \tag{2.8}$$

The function is utilized primarily in connection with the following important property typically (but not always) satisfied by differentials.

CONDITION L.

$$T(F; \Delta) = \int_{-\infty}^{\infty} T[F; x]d\Delta(x), \quad \Delta \in \mathcal{D}(F). \ \square \tag{2.9}$$

An important implication of Condition L, obtained by the substitution $\Delta = \delta_x - F$ in (2.9), is

$$T[F; x] = T(F; \delta_x - F), \quad -\infty < x < \infty, \tag{2.10a}$$

or equivalently

$$\mu(T, F) = 0. \tag{2.10b}$$

(Besides containing (2.10), Condition L represents a strengthening of the *linearity* property of the differential. Indeed, by the usual linearity expressed by (2.3), it follows easily that

$$T(F;\, G - F) = \int_{-\infty}^{\infty} T[F;\, x]\,d[G(x) - F(x)] + \mu(T,\, F)$$

for the case of G discrete with finite support. Also, by linearity again, $T(F;\, G - H) = T(F;\, G - F) - T(F;\, H - F)$. Thus, under (2.10), we have (2.9) for all $\Delta = c(G - H)$, where c is constant and G and H are discrete with finite support. Hence Condition L merely extends (2.9), in the presence of (2.10), to general G and H.)

Now note that when the result of Lemma 2.1 is applicable for G given by δ_x, each x, then Condition L may be written in the form

$$T(F;\, \Delta) = \int_{-\infty}^{\infty} D_{\delta_x} T(F)\,d\Delta(x), \quad \Delta \in \mathcal{D}(F). \tag{2.11}$$

This suggests a convenient scheme for determination of the differential. First obtain, by routine calculus methods of differentiation,

$$D_{\delta_x} T(F) = \lim_{t \to 0+} \frac{T(F + t(\delta_x - F)) - T(F)}{t}, \quad -\infty < x < \infty. \tag{2.12}$$

Then, motivated by the anticipated validity of (2.11), adopt it as the definition of $T(F;\, \Delta)$. Finally, establish that $T(F;\, \Delta)$ so defined fulfills the definition of the differential. The difficult part is the latter step. The following lemma characterizes the scheme.

LEMMA 2.2. *Suppose that* (2.1) *holds for* $T(F;\, \Delta)$ *defined by* (2.11). *Then* $T(F;\, \Delta)$ *is the differential of* T *at* F *with respect to* $||\cdot||$ *and* G_F, *and Condition* L *holds. Also,*

$$T[F;\, x] = D_{\delta_x} T(F) - \int_{-\infty}^{\infty} D_{\delta_x} T(F)\,dF(x), \quad -\infty < x < \infty. \tag{2.13}$$

If, further, G_F is weakly starshaped at F with respect to F, then

(2.14a)
$$\int_{-\infty}^{\infty} D_{\delta_x} T(F)dF(x) = 0,$$

and hence

(2.14b)
$$T[F; x] = D_{\delta_x} T(F), \quad -\infty < x < \infty.$$

PROOF. Clearly $T(F; \Delta)$ defined by (2.11) is linear in the argument Δ. Thus fulfillment of (2.1) completes the requirements for $T(F; \Delta)$ to be the differential of T. Putting $\Delta = \delta_x - F$ in (2.11), we obtain

(2.15)
$$T(F; \delta_x - F) = D_{\delta_x} T(F) - \int_{-\infty}^{\infty} D_{\delta_x} T(F)dF(x), \ -\infty < x < \infty,$$

which implies (2.10) and hence (2.13). From (2.11) and (2.13), Condition L follows. The final claim of the lemma is straightforward from (2.15) and Lemma 2.1. □

2.2. *The statistical role of the differential.* Consider a sample of (not necessarily independent) observations $X_1, \ldots, X_n$ from a distribution F. Let us denote by F_n the sample distribution function based on $X_1, \ldots, X_n$, i.e.,

(2.16)
$$F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} .$$

Many statistics of interest may be represented either exactly or approximately as "statistical functions" $T(F_n)$ for some functional T. In the case that T possesses a differential, the analysis of $T(F_n)$ relative to the associated "parameter" $T(F)$ may be reduced via the differential to a problem of standard form in terms of a *sum* of random variables. Specifically, under appropriate assumptions entailing the convergence of F_n to F (see Lemmas 2.5 and 2.7), the asymptotic behavior of $T(F_n) - T(F)$ corresponds closely to that of $T(F; F_n - F)$. But, by linearity of the differential, the latter random variable may be written as an average of certain special random variables, i.e.,

(2.17a) $$T(F;\ F_n - F) = \frac{1}{n}\sum_{i=1}^{n} T(F;\ \delta_{X_i} - F)$$

(2.17b) $$= \frac{1}{n}\sum_{i=1}^{n} T[F;\ X_i] + \mu(T,\ F).$$

From this representation, and through appeal to central limit theory and other probabilistic results for *sums* of random variables, key results for $T(F_n)$ follow. The asymptotic *distribution theory* (normality) of $T(F_n) - T(F)$ will be given in Theorem 2.1. The *almost sure behavior* will be characterized by a *law of iterated logarithm* given in Theorem 2.2.

As a further statistical application of the differential, the function $T[F;\ x]$ may be regarded as an "influence curve." For, by (2.17), the *error of approximation* involved in estimating $T(F)$ by $T(F_n)$ is given approximately by

$$\frac{1}{n}\sum_{i=1}^{n} T[F;\ X_i] + \mu(T,\ F).$$

Thus $T[F;\ X_i]$ measures the approximate "influence" of the observation X_i toward the error $T(F_n) - T(F)$. This notion is due to Hampel (1968), (1974), who defines

$$\Omega_{T,F}(x) = D_{\delta_x} T(F), \quad -\infty < x < \infty,$$

as the *influence curve* of the estimator $T(F_n)$ for $T(F)$. (Recall that in the presence of Condition L and the condition that $\mathcal{G}_F$ is weakly starshaped at F with respect to F, we have $T[F;\ x] = D_{\delta_x} T(F)$.) A number of characteristics of the function $\Omega_{T,F}(\cdot)$ are interpreted by Hampel as measures of robustness properties of the estimator $T(F_n)$.

We may interpret $\mu(T,\ F)$ as an asymptotic *bias* quantity. In typical applications, $\mu(T,\ F) = 0$.

2.3. *A quasi-differential.* Before pursuing formally these statistical results based on the "differential approach," we introduce a modification of the method designed to broaden the scope

of its statistical application. From the foregoing discussion concerning approximation of $T(F_n) - T(F)$ by $T(F; F_n - F)$, it is clear that we would be equally well served by an approximation of the form

$$T_F(F_n) \circ T(F; F_n - F),$$

where $T_F(\cdot)$ is an auxiliary functional such that $T_F(F_n)$ converges to $T_F(F)$ in an appropriate stochastic sense. This motivates us to allow $T(F; G - F)$ to be replaced, in condition (2.1) in the definition of the differential, by a quantity of the more general form $T_F(G) \circ T(F; G - F)$. Accordingly, we introduce the following notion (*not* the same as the "quasi-differential" of Dieudonne (1960), p. 151):

DEFINITION. We say that a functional T defined on $\mathcal{F}$ has a *quasi-differential* at the point of $F \in \mathcal{F}$ with respect to the norm $||\cdot||$, the set $\mathcal{G}_F \subset \mathcal{F}$, and the functional $T_F(\cdot)$ defined on $\mathcal{G}_F$, if there exists a quantity $T(F; \Delta)$, defined on $\Delta \in \mathcal{D}(\mathcal{F})$, which is linear in the argument Δ and satisfies the condition

$$\lim_{\substack{||G-F||\to 0 \\ G \in \mathcal{G}_F}} \frac{T(G) - T(F) - T_F(G)\, T(F; G - F)}{||G - F||} = 0. \tag{2.18a}$$

Without loss of generality we also assume

$$T_F(F) = 1. \ \square \tag{2.18b}$$

($T(F; \Delta)$ is called the "quasi-differential.")

In the case that $T_F(G) \equiv 1$, the quasi-differential is in fact the differential. In other cases, the quasi-differential can also in fact be a differential, but the stronger version may be more difficult to verify or may entail additional restrictions on the structure of $T(\cdot)$. In such cases, we have the option of following the easier route and nevertheless achieving a satisfactory approximation to $T(F_n) - T(F)$.

For the purpose of determining the quasi-differential, the scheme based on (2.11) for finding the differential is again applicable. The following analogue of Lemma 2.2 is easily checked.

LEMMA 2.3. *Suppose that* (2.18) *holds for* $T(F;\Delta)$ *defined by* (2.11). *Then* $T(F;\Delta)$ *is the quasi-differential of* T *at* F *with respect to* $||\cdot||$, G_F *and* $T_F(\cdot)$, *and Condition* L *holds. Also,* (2.13) *holds. If, further,* G_F *is weakly starshaped at* F *with respect to* $\mathcal{F}$, *then* (2.14) *holds.*

2.4. *Stochastic quasi-differentials.* Although the preceding extension of the concept of differential is adequate for many statistical applications, further broadening is called for in some situations. We require merely that T possess a quasi-differential in a suitable *stochastic* sense. Specifically, (2.18a) is required to hold only for the sequence $\{F_n\}$. We thus introduce

DEFINITION. We say that a functional T defined on $\mathcal{F}$ has a *weak stochastic quasi-differential* at the point $F \in \mathcal{F}$ with respect to the norm $||\cdot||$, the sequence of random variables $\{X_i\}$, and the functional $T_F(\cdot)$, if $T_F(F_n)$ is always defined and

$$(2.19) \qquad ||F_n - F|| \xrightarrow{p} 0, \; n \to \infty,$$

and there exists a quantity $T(F;\Delta)$, defined on $\Delta \in \mathcal{D}(\mathcal{F})$, which is linear in the argument Δ and satisfies the condition

$$(2.20) \qquad \frac{T(F_n) - T(F) - T_F(F_n)\, T(F; F_n - F)}{||F_n - F||} \xrightarrow{p} 0, \; n \to \infty.$$

Without loss of generality we also assume (2.18b). □ ($T(F;\Delta)$ is called the "stochastic quasi-differential.") Note that the observations $\{X_i\}$ are not required to be independent.

DEFINITION. If the convergences in (2.19) and (2.20) hold *with probability* 1, then we say that T has a *strong stochastic quasi-differential.* □

Note that the choice of norm $||\cdot||$ must serve two somewhat conflicting purposes, corresponding to conditions (2.19) and (2.20). The first is more easily satisfied for $||\cdot||$ "small," whereas the other is more easily satisfied if $||\cdot||$ is "large."

We now examine the connections between the strict quasi-differential and its stochastic versions. For this purpose, put

$$L(G, F) = \frac{T(G) - T(F) - T_F(G)\, T(F;\, G - F)}{||G - F||}, \text{ if } ||G - F|| > 0,$$

$$(2.21) \qquad = 0, \text{ if } ||G - F|| = 0,$$

when the relevant quantities are defined.

LEMMA 2.4. *Suppose that* T *has a quasi-differential at* F *with respect to* $||\cdot||$, G_F *and* $T_F(\cdot)$.

(i) *If*

$$(2.22a) \qquad P\{F_n \in G_F\} \to 1,\ n \to \infty,$$

and

$$(2.22b) \qquad ||F_n - F|| \xrightarrow{p} 0,\ n \to \infty,$$

then

$$(2.22c) \qquad L(F_n, F) \xrightarrow{p} 0,\ n \to \infty.$$

(ii) *If*

$$(2.23a) \qquad P\{F_n \in G_F, \textit{ all } n \textit{ sufficiently large}\} = 1$$

and

$$(2.23b) \qquad ||F_n - F|| \xrightarrow{w.p.1} 0,\ n \to \infty,$$

then

(2.23c) $$L(F_n, F) \xrightarrow{\text{w.p.1}} 0, \quad n \to \infty.$$

PROOF. (i) Let $\varepsilon > 0$ be given. By hypothesis, there exists $\delta > 0$ such that $||G - F|| < \delta$, $G \in \mathcal{G}_F$, implies $|L(G, F)| < \varepsilon$. Then

(2.24) $$P\{|L(F_n, F)| > \varepsilon\} \le P\{||F_n - F|| \ge \delta\} + P\{F_n \notin \mathcal{G}_F\}.$$

By (2.22a) and (2.22b), the right-hand side of (2.24) tends to 0 as $n \to \infty$. Thus (2.22c) follows.

(ii) Trivial. □

Under the conditions of the lemma, T has a *weak* stochastic quasi-differential in case (i), a *strong* stochastic quasi-differential in case (ii).

The scheme previously discussed for finding the form of the quasi-differential serves also in seeking a stochastic version.

It is interesting and useful that the concept of differential may also be formulated even when no functional is explicitly involved.

DEFINITION. We say that a sequence of statistics $\{T_n\}$, $T_n = T_n(X_1, \ldots, X_n)$, satisfying

(2.25a) $$T_n \xrightarrow{p} T_o, \quad n \to \infty,$$

has a *weak stochastic quasi-differential* at (T_o, F) with respect to the norm $||\cdot||$, the sequence of random variables $\{X_i\}$, and the sequence of random variables $\{Z_n\}$, $Z_n = Z_n(X_1, \ldots, X_n)$, if

(2.25b) $$||F_n - F|| \xrightarrow{p} 0, \quad n \to \infty,$$

and there exists a quantity $T(F; \Delta)$, defined on $\Delta \in \mathcal{D}(F)$, which is linear in the argument Δ and satisfies the condition

(2.25c) $$\frac{T_n - T_o - Z_n T(F; F_n - F)}{||F_n - F||} \xrightarrow{p} 0, \quad n \to \infty. \ \square$$

DEFINITION. If the convergences in (2.25) hold *with probability* 1, then we say that $\{T_n\}$ has a *strong stochastic quasi-differential*. □

In the remainder of this section, we shall confine attention to the versions based on *functionals*. However, in Section 4 some application of the preceding generalization will be noted.

2.5. *Statistical applications of stochastic quasi-differentials*. The role of the stochastic quasi-differential in obtaining *asymptotic normality* of statistical functions is seen precisely from the following result, whose proof is immediate.

LEMMA 2.5. *Let* $\{X_i\}$ *be a sequence of observations (not necessarily independent) on a distribution* F. *Let* T *be a functional defined on a (convex) set* $\mathcal{F}$ *containing* F. *Suppose that* T *has a weak stochastic quasi-differential at* F *with respect to* $||\cdot||$, $\{X_i\}$ *and* $T_F(\cdot)$. *Suppose further that*

$$\sqrt{n}\ ||F_n - F|| = O_p(1),\ n \to \infty. \tag{2.26}$$

Then,

$$\sqrt{n}\ |T(F_n) - T(F) - T_F(F_n)\, T(F; F_n - F)| \xrightarrow{p} 0,\ n \to \infty. \tag{2.27}$$

In particular, we now make application of Lemma 2.5 for the case of *independent* observations and with respect to the norm

$$||h||_\infty = \sup_{-\infty<x<\infty} |h(x)|\ .$$

We shall utilize the following result.

LEMMA 2.6. *Let* $\{X_i\}$ *be a sequence of independent observations on a nondegenerate distribution* F. *Then*

$$\sqrt{n}\ ||F_n - F||_\infty \xrightarrow{d} Z_F\ , \tag{2.28}$$

where Z_F *is positive with probability* 1.

The proof of (2.28) for the case of F continuous was first given by Kolmogorov (1933). The distribution of Z_F was given explicitly and seen not to depend upon F. Extension to the case of F having finitely many discontinuities and not being purely atomic was obtained by Schmid (1958). Here also the distribution of Z_F was given explicitly; it depends upon F in the case of discontinuities. The general case is treated in Billingsley (1968), Section 16. By his Theorem 16.4 and subsequent discussion, it is seen that the random variable Z_F may be characterized as

(2.29) $$Z_F = \sup_{-\infty<x<\infty} |W^{\circ}(F(x))| ,$$

where W° is the "tied-down Wiener process" (or "Brownian Bridge"), i.e., the Gaussian stochastic process $\{W^{\circ}(t),\ 0 \le t \le 1\}$ specified by $E\{W^{\circ}(t)\} = 0$ and $E\{W(s)\ W(t)\} = s(1 - t),\ 0 \le s \le t \le 1$. It is evident from (2.29) that Z_F is positive with probability 1, except in the case that F is degenerate.

(For completeness we make explicit the line of development in Billingsley (1968) leading to (2.28) and (2.29). We may represent the X_i's as $X_i = F^{-1}(U_i)$, where $\{U_i\}$ is a sequence of independent uniform random variables on [0,1]. Let G_n be the sample distribution function of $U_1, \ldots, U_n$. Then $F_n(x) = G_n(F(x))$, $-\infty<x<\infty$, so that

$$\sqrt{n}\ ||F_n - F||_\infty = \sqrt{n} \sup_{-\infty<x<\infty} |G_n(F(x)) - F(x)| = \sqrt{n} \sup_{t\epsilon A} |G_n(t) - t|,$$

where $A = \{F(x): -\infty<x<\infty\}$. For the empirical stochastic process $Y_n(t) = \sqrt{n}[G_n(t) - t]$, $0 \le t \le 1$, Billingsley (Theorem 16.4) establishes that Y_n converges in distribution to W°. This weak convergence is established in the space D of functions $x(\cdot)$ on [0,1] that are right-continuous and have left-hand limits. Consider the mapping $h(x) = \sup_{t\epsilon A}|x(t)|$ of D into the real line. We have

$$\sqrt{n}\ ||F_n - F||_\infty = h(Y_n).$$

Now, with respect to the Skorohod topology in D, the mapping h is seen to be continuous at points x belonging to the space C of continuous functions on [0,1]. For, if x_n converges to x in the Skorohod topology, and $x \in C$, then x_n converges to x in the uniform topology, and from this it follows that $h(x_n)$ converges to $h(x)$. Thus h is continuous with probability 1 under the measure corresponding to W^o. Hence, by Billingsley's Theorems 5·1 and 16.4, $h(Y_n)$ converges in distribution to $h(W^o) = Z_F$.)

Also, we shall be concerned, through the representation (2.17), with the random variables $T[F; X_i]$, $1 \le i \le n$, and with the quantity $\mu(T, F)$. We may express (2.7) in the form

$$\mu(T, F) = E_F\{T(F; \delta_X - F)\}. \tag{2.30}$$

By (2.8), we have $E_F\{T[F; X]\} = 0$. Let us put

$$\sigma^2(T, F) = \mathrm{Var}_F\{T[F; X]\}. \tag{2.31}$$

THEOREM 2.1. *Let* $\{X_i\}$ *be a sequence of independent observations on a distribution* F. *Let* T *be a functional defined on a (convex) set* $\mathcal{F}$ *containing* F. *Suppose that* T *has a weak stochastic quasi-differential at* F *with respect to* $||\cdot||_\infty$, $\{X_i\}$ *and* $T_F(\cdot)$. *Assume that* $0 < \sigma^2(T, F) < \infty$. *Further, assume that*

(i) *if* $\mu(T, F) = 0$, *then*

$$T_F(F_n) \xrightarrow{p} 1,\ n \to \infty; \tag{2.32}$$

(ii) *if* $\mu(T, F) \neq 0$, *then*

$$\sqrt{n}\,[T_F(F_n) - 1] \xrightarrow{p} 0,\ n \to \infty. \tag{2.33}$$

Then

$$\sqrt{n}\,[T(F_n) - T(F) - \mu(T, F)] \xrightarrow{d} N(0, \sigma^2(T, F)). \tag{2.34}$$

PROOF. By the independence assumption, Lemma 2.6 is applicable and thus (2.26) holds for the norm $||\cdot||_\infty$. Thus, in turn, Lemma 2.5 is applicable and yields a reduction to the random variable

$$(2.35) \qquad \sqrt{n}\, T_F(F_n)\, [T(F; F_n - F) - \mu(T, F)] + \mu(T, F) \sqrt{n}\, [T_F(F_n) - 1]$$

in lieu of the random variable in (2.34). Utilizing the representation (2.17), with either (2.32) or (2.33) as appropriate, we have by routine application of central limit theory that the random variable in (2.35) converges in distribution to $N(0, \sigma^2(T, F))$. □

In typical applications, the asymptotic bias quantity $\mu(T, F)$ is 0 and we need merely to verify (2.32) instead of the stronger condition (2.33). Of course, in many applications $T_F(G) \equiv 1$, in which case (2.32) and (2.33) hold trivially.

Under a condition on F_n slightly different from (2.26), but closely related to the latter, we may characterize the *almost sure behavior* of $T(F_n)$. The role of the stochastic quasi-differential is exhibited in the following result, whose proof is immediate.

LEMMA 2.7. *Let* $\{X_i\}$ *be a sequence of observations (not necessarily independent) on a distribution* F. *Let* T *be a functional defined on a (convex) set* $\mathcal{F}$ *containing* F. *Suppose that* T *has a strong stochastic quasi-differential at* F *with respect to* $||\cdot||$, $\{X_i\}$ *and* $T_F(\cdot)$. *Suppose further that*

$$(2.36) \qquad \sqrt{n}\, ||F_n - F|| = O(\sqrt{\log \log n}),\ n \to \infty,\ w.p.\ 1.$$

Then

$$(2.37) \qquad \frac{\sqrt{n}\, |T(F_n) - T(F) - T_F(F_n)\, T(F; F_n - F)|}{\sqrt{\log \log n}} \xrightarrow{wp1} 0,\ n \to \infty.$$

In particular, similar to our application of Lemma 2.5, we now consider the

case of *independent* observations and the norm $||\cdot||_\infty$. We shall utilize the following result.

LEMMA 2.8. *Let* $\{X_i\}$ *be a sequence of independent observations on a distribution* F. *Then*

$$\overline{\lim_{n\to\infty}} \frac{\sqrt{n}\,||F_n - F||_\infty}{\sqrt{\log\log n}} = \sup_{-\infty<x<\infty} \sqrt{2F(t)[1 - F(t)]} \quad w.p.\ 1. \tag{2.38}$$

The proof of (2.38) in the case of F continuous was given by Chung (1949). Extension to the case of F having discontinuities is due to Richter (1974).

We now establish a *law of iterated logarithm* for statistical functions.

THEOREM 2.2. *Let* $\{X_i\}$ *be a sequence of independent observations on a distribution* F. *Let* T *be a functional defined on a (convex) set* $\mathcal{F}$ *containing* F. *Suppose that* T *has a strong stochastic quasi-differential at* F *with respect to* $||\cdot||_\infty$, $\{X_i\}$ *and* $T_F(\cdot)$. *Assume that* $0 < \sigma^2(T, F) < \infty$. *Further, assume that*

(i) *if* $\mu(T, F) = 0$, *then*

$$T_F(F_n) \xrightarrow{wp1} 1,\ n \to \infty; \tag{2.39}$$

(ii) *if* $\mu(T, F) \neq 0$, *then*

$$\frac{\sqrt{n}\,[T_F(F_n) - 1]}{\sqrt{\log\log n}} \xrightarrow{wp1} 0,\ n \to \infty. \tag{2.40}$$

Then

$$\overline{\lim_{n\to\infty}} \frac{\sqrt{n}\,[T(F_n) - T(F) - \mu(T, F)]}{\sqrt{2\,\sigma^2(T, F)\log\log n}} = 1 \quad w.p.1\ . \tag{2.41}$$

PROOF. By the independence assumption and Lemma 2.8, we have (2.36) for the norm $||\cdot||_\infty$. It follows by Lemma 2.7 that

$$\xi_{1n} - \xi_{2n} - \xi_{3n} \xrightarrow{wp1} 0,\ n \to \infty, \tag{2.42}$$

where

$$\xi_{1n} = \frac{\sqrt{n}\,[T(F_n) - T(F) - \mu(T, F)]}{\sqrt{2\,\sigma^2(T, F)\log\log n}},$$

$$\xi_{2n} = T_F(F_n)\,\frac{\sqrt{n}\,[T(F; F_n - F) - \mu(T, F)]}{\sqrt{2\,\sigma^2(T, F)\log\log n}},$$

and

$$\xi_{3n} = \mu(T, F)\,\frac{\sqrt{n}[T_F(F_n) - 1]}{\sqrt{2\,\sigma^2(T, F)\log\log n}}.$$

Utilizing the representation (2.17) and the classical law of the iterated logarithm of Hartman and Wintner (1941), we have

$$\varlimsup_{n\to\infty} \frac{\sqrt{n}\,[T(F; F_n - F) - \mu(T, F)]}{\sqrt{2\,\sigma^2(T, F)\log\log n}} = 1 \quad \text{w.p.1.} \tag{2.43}$$

By (2.39), which holds whether $\mu(T, F) = 0$ or $\neq 0$, we thus have

$$\varlimsup_{n\to\infty} \xi_{2n} = 1 \quad \text{w.p.1.} \tag{2.44}$$

By (2.40),

$$\xi_{3n} \xrightarrow{\text{wpl}} 0,\ n \to \infty. \tag{2.45}$$

By (2.42), (2.44) and (2.45), we readily obtain

$$\varlimsup_{n\to\infty} \xi_{1n} = 1 \quad \text{w.p.1.},$$

i.e., (2.41) holds. □

2.6. *Illustration: the variance parameter.* The variance parameter of a distribution F may be represented as a *functional* as follows:

$$(2.46) \qquad T(F) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x_1, x_2)\, dF(x_1)\, dF(x_2),$$

where $h(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2 - 2x_1x_2)$. In this case, evaluation of $T(\cdot)$ at the sample distribution F_n yields

$$T(F_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

the "sample variance."

In order to explore the possibility of a differential for $T(\cdot)$, we begin with (2.4) and readily obtain

$$(2.47) \qquad D_G\, T(F) = 2\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)\, dF(y)\, d[G(x) - F(x)]$$

and thus

$$(2.48) \qquad D_{\delta_x}\, T(F) = 2\left[\int_{-\infty}^{\infty} h(x, y)\, dF(y) - T(F)\right],$$

or, after some reduction,

$$(2.49) \qquad D_{\delta_x}\, T(F) = (x - \mu_F)^2 - \sigma_F^2,$$

where μ_F and σ_F^2 denote the mean and variance of F. Thus, defining a "candidate" differential for T in correspondence with (2.11), we have

$$T(F;\, \Delta) = \int_{-\infty}^{\infty} (x - \mu_F)^2\, d\Delta(x) - \sigma_F^2 \int_{-\infty}^{\infty} d\Delta(x)$$

and, in particular,

$$(2.50) \qquad T(F;G-F) = \int_{-\infty}^{\infty} (x - \mu_F)^2 dG(x) - \sigma_F^2 = \int_{-\infty}^{\infty} x^2 d[G(x)-F(x)] - 2\mu_G\mu_F + 2\mu_F^2,$$

where μ_G denotes the mean of G. Also,

$$(2.51) \qquad T(G) - T(F) = \int_{-\infty}^{\infty} x^2 d[G(x) - F(x)] - \mu_G^{\,2} + \mu_F^{\,2}.$$

Hence

$$(2.52) \qquad T(G) - T(F) - T(F;\, G - F) = -(\mu_G - \mu_F)^2.$$

In order to establish that $T(F;\, \Delta)$ is a differential of T with respect to $||\cdot||_\infty$ and a set $G_F \subset \mathbf{F}$, we need to show that

$$(2.53) \qquad \lim_{\substack{||G - F||_\infty \to 0 \\ G \in G_F}} L(G,\, F) = 0,$$

where

$$(2.54) \qquad L(G,\, F) = \frac{T(G) - T(F) - T(F;\, G - F)}{||G - F||_\infty} = \frac{-[\int_{-\infty}^{\infty} x d(G - F)]^2}{||G - F||_\infty}.$$

Unfortunately, it is found by considering specific examples that in general $L(G,\, F)$ need not $\to 0$ as $||G - F||_\infty \to 0$. However, we are able relatively easily to establish a stochastic version of (2.53). Write

$$(2.55) \qquad L(F_n,\, F) = -\left[\sqrt{n}\, \frac{\sum_{i=1}^{n} (X_i - \mu_F)}{n}\right] \cdot \frac{1}{\sqrt{n}\, ||F_n - F||_\infty} \cdot \frac{\sum_{i=1}^{n} (X_i - \mu_F)}{n}.$$

Assume that the X_i's are I.I.D. By the central limit theorem, the first factor in (2.55) converges in distribution to a finite-valued random variable. By Lemma 2.6 and the convergence (2.28),

$$(2.56) \qquad \frac{1}{\sqrt{n}\, ||F_n - F||_\infty} \xrightarrow{d} \frac{1}{Z_F},$$

since the function $g(x) = x^{-1}$ is continuous with probability 1 with respect to the distribution of Z_F. Thus the second factor in (2.55) also converges in distribution to a finite-valued random variable. Finally, by the strong law of

large numbers, the third factor in (2.55) converges to 0 with probability 1. It follows that

$$L(F_n, F) \xrightarrow{p} 0, \tag{2.57}$$

i.e., that $T(F; \Delta)$ is a weak stochastic differential for $T(\cdot)$ at F with respect to $||\cdot||_\infty$ and $\{X_i\}$. By an extended version of Lemma 2.2, we have that (2.13) holds. In conjunction with (2.6) and (2.49), we thus have

$$T[F; x] = (x - \mu_F)^2 - \sigma_F^{\ 2}\ , \ -\infty < x < \infty,$$

and accordingly

$$\mu(T, F) = 0$$

and

$$\sigma^2(T, F) = \mathrm{Var}_F\{(X - \mu_F)^2\} = \mu_4(F) - \sigma_F^{\ 4},$$

where $\mu_4(F) = E_F\{(X - \mu_F)^4\}$, the 4-th central moment of F. It thus follows from Theorem 2.1 that

$$\sqrt{n}\ [T(F_n) - T(F)] \xrightarrow{d} N(0,\ \mu_4(F) - \sigma_F^{\ 4}),$$

a well-known result expressing asymptotic normality of the sample variance.

If we now seek to establish a law of iterated logarithm for $T(F_n)$ via Theorem 2.2, we need to establish that $L(F_n, F) \xrightarrow{wpl} 0$. In place of (2.55), we write

$$L(F_n, F) = -\left|\frac{\sqrt{n}\ \sum_{i=1}^{n} (X_i - \mu_F)}{\sqrt{\log \log n}}\right|^2 \cdot \frac{\log \log n}{n\ ||F_n - F||_\infty}\ . \tag{2.58}$$

The first factor is O(1) with probability 1, by the classical law of the iterated logarithm of Hartman and Wintner (1941). Thus it suffices to show that

$$(2.59) \qquad \frac{\log \log n}{\sqrt{n}\, ||F_n - F||_\infty} \xrightarrow{wpl} 0.$$

However, this does not appear to be an immediate corollary of any known theorem on the behavior of $\sqrt{n}\, ||F_n - F||_\infty$. Nevertheless we conjecture (2.59) to be true and leave it to be dealt with elsewhere. In any case, the desired law of iterated logarithm for the sample variance does indeed hold. One may obtain it by a direct argument starting with (2.52), with $G = F_n$.

A general class of statistics may be handled similarly to the foregoing treatment. These are the statistical functions corresponding to functionals of the form

$$(2.60) \qquad T(F) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \cdots, x_c)\, dF(x_1) \cdots dF(x_c).$$

The related statistical functions may be written as

$$T(F_n) = \frac{1}{n^c} \sum_{i_1=1}^{n} \cdots \sum_{i_c=1}^{n} h(X_{i_1}, \cdots, X_{i_c}).$$

We note that they are closely related to the U-statistics of Hoeffding (1948), wherein asymptotic normality is established. The law of the iterated logarithm for this class of statistic is noted in Serfling (1971).

2.7. *Norms of interest.* The norm

$$(2.61) \qquad ||h||_\infty = \sup_{-\infty<x<\infty} |h(x)|$$

receives special attention primarily because of the many useful results available for the random variable $||F_n - F||_\infty$, as seen for example in Lemmas 2.6 and 2.8. Noting, however, the conflicting demands imposed on any norm by the conditions (2.19) and (2.20), we consider other possibilities also.

A useful alternative norm is given by

(2.62) $$||h||_q = \left|\left|\frac{h}{q}\right|\right|_\infty = \sup_{-\infty<x<\infty}\left|\frac{h(x)}{q(x)}\right|,$$

where $q(\cdot)$ is a given bounded and continuous function, usually satisfying restrictions on the rate of convergence to 0 as $x \to \pm\infty$. For use in the form $||F_n - F||_q$, the choice of $q(\cdot)$ usually depends upon F. Effective utilization of this norm has been made by Gregory (1976) and Boos (1977a). Indeed, versions of Theorem 2.1 and 2.2 are available with $||\cdot||_\infty$ replaced by $||\cdot||_q$ for q in a particular class.

Another norm of interest is the *variation* norm,

(2.63) $$||h||_V = \lim_{\substack{a\to-\infty\\ b\to+\infty}} V_{a,b}(h),$$

where

$$V_{a,b}(h) = \sup \sum_{i=1}^{k} |h(x_i) - h(x_{i-1})|,$$

the supremum being taken over all partitions $a = x_0 < x_1 < \cdots < x_k = b$ of the interval [a, b]. Except in special cases, the norm $||F_n - F||_V$ is not effective for our purposes. However, in Section 4, we shall utilize $||h||_V$ in a different way.

2.8. *Concluding remarks.* (i) Clearly, the concepts presented here may be extended to the case of differentation of higher order.

(ii) The "mechanical" aspect of the differential approach is worthy of note. For example, in connection with the asymptotic normality result Theorem 2.1, the variance parameter $\sigma^2(T, F)$ may be obtained mechanically without having yet established actual existence of the differential.

(iii) The sequence of sample distribution functions $\{F_n\}$ possesses a unique effectiveness among sequences converging weakly to F. This has been illustrated in connection with the sample variance, for which a *stochastic* differential may be established in the absence of a nonstochastic version. Similarly, in Section 4, a

stochastic quasi-differential will be established for a wider class of M-functionals than in the nonstochastic case.

(iv) The "differential approach" was introduced into the statistical context by von Mises (1947). Rather than employing a differential, he developed the Taylor expansion of $T(F + t(F_n - F))$, considered as a function of the real variable t, $0 \le t \le 1$, and treated the remainder term as asymptotically negligible in appropriate senses. This also leads to (2.27). The extended treatment by Filippova (1962) is in the same vein. However, the development by Kallianpur and Rao (1955) is carried out in the setting of Frechet differentiation, with special reference to the norm $||\cdot||_\infty$. Gregory (1976) provides a fuller development in this context, with special reference to the norms of type $||\cdot||_q$, and applies his theory to obtain asymptotic normality results for a wide class of linear functions of order statistics, under the assumption of absolute continuity of F. By the use of the broader versions of differential technique which we have developed in the present section, a wider class of linear functions of order statistics may be treated and the continuity of F may be relaxed, and almost sure behavior may be characterized as well as asymptotic normality. See Boos (1977a,b) for full development. Finally, we mention the recent work of Beran (1977a,b) developing and applying a theory of Frechet differentiation of functionals with respect to the Hellinger metric.

(v) Extension of our treatment to the case of *dependent* observations X_i on F is straightforward. All that is needed is a handling of $\sum_1^n T[F; X_i]$ and of $||F_n - F||_\sigma$ for such cases.

(vi) Extension to the *multi-dimensional* case is possible also. For example, the condition that $\sqrt{n}\, ||F_n - F||_\infty = O_p(1)$, $n \to \infty$, needed for an extension of Lemma 2.5, follows for F a distribution on R^k by a result of Kiefer and Wolfowitz (1958).

(vii) Throughout it is assumed implicitly that $T(F; F_n - F)$ is measurable.

(viii) Earlier we noted that condition L implies (2.10b), i.e., $\mu(T, F) = 0$. This is also implied by the condition that F be discrete with finite support.

3. A useful inequality. The endeavor to establish a (stochastic quasi-) differential often entails a quantity of the form

$$\int h\,d(G - F),$$

where F and G are distribution functions. For example, refer to (2.54) in our treatment of the variance functional. In some cases the following inequality is quite useful. We shall exploit it in Section 4. (See Section 2.7 for the definitions of the norms.)

LEMMA 3.1. *Let the function* h *be continuous and of bounded variation on* R. *Let* G *and* F *be distribution functions. Then*

$$|\int h\,d(G - F)| \leq ||h||_V \; ||G - F||_\infty. \tag{3.1}$$

The proof is based on the following two results. The first is easily proved (or see Natanson (1961), p. 232). The second is given by Dunford and Schwartz (1958), p. 154.

LEMMA 3.2. *Suppose that* f *is bounded on* R, g *is of bounded variation on* R, *and* $\int f\,dg$ *exists. Then*

$$|\int f\,dg| \leq ||f||_\infty \cdot ||g||_V. \tag{3.2}$$

LEMMA 3.3. *Let* f *and* g *be of bounded variation on an interval* (a, b), *allowing* $a = -\infty$ *and* $b = +\infty$. *Suppose that one of the functions is continuous in* (a, b) *and that the other is continuous on the right. Then*

$$\int_a^b f(x)\,dg(x) + \int_a^b g(x)\,df(x) = f(b-)\,g(b-) - f(a+)\,g(a+). \tag{3.3}$$

PROOF OF LEMMA 3.1. By Lemma 3.3,

$$\int h\,d(G - F) = -\int (G - F)\,dh + h(\infty)(G - F)(\infty) - h(-\infty)(G - F)(-\infty).$$

Since $|h(\pm\infty)| < \infty$, we have $h(\infty)(G - F)(\infty) = h(-\infty)(G - F)(-\infty) = 0$. By Lemma 3.2, we obtain the desired result. □

4. Application to M-estimation of location parameters.

4.0. *Preliminary remarks.* The setting for "M-estimation" of a location parameter is based on a sample $X_1,\ldots,X_n$ from a distribution F and a function ψ such that the solution T of the equation

$$(4.1) \qquad \int \psi(x - T)dF(x) = 0$$

coincides with the parameter in question. In this situation, a natural estimator of the parameter is provided by the solution T_n of the analogous equation based on the sample distribution function F_n in place of F, i.e., the equation

$$(4.2) \qquad \int \psi(x - T_n)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} \psi(X_i - T_n) = 0.$$

In this section we first examine some particular ψ functions of interest in such estimation problems. Then we give a precise formulation of the problem, whereby the solution T of (4.1) becomes a well-defined *functional* T(F) defined on distributions F, in which case the solution T_n of (4.2) is just $T(F_n)$. The formulation will take into account the scope of types of ψ function of practical interest. We then characterize the behavior of $T(F_n)$ via the methodology of Section 2. As a preliminary step, the behavior of $T(G_n)$ for weakly convergent sequences $\{G_n\}$ is explored. This yields results on the strong consistency of $T(F_n)$. Then we establish that, under very broad assumptions regarding ψ and F, the functional T possesses a strong stochastic quasi-differential. Conditions for a strict quasi-differential are also given. Using the differentiability results, we obtain asymptotic normality and the law of the iterated logarithm for $T(F_n)$. Following this development, we briefly consider some specific examples. We then proceed to extend the results to one-step versions and to the case of scale unknown. Finally, we make comparisons with related results in the literature.

4.1. *Varieties of ψ function.* Different choices of ψ lead to different estimators. For example, the function $\psi(x) = x$ yields the *sample mean*, $\bar{X}$. The function $\psi(x) = \mathrm{sgn}(x)$ yields the *sample median.*

Classical *maximum likelihood estimation* corresponds to the assumption that F is of the form $F_o(x - \theta)$, where θ is the unknown location parameter and F_o is a specified known distribution function. If F_o has a differentiable density f_o and ψ is given by

$$\psi(x) = -\frac{f_o'(x)}{f_o(x)}, \quad -\infty < x < \infty, \tag{4.3}$$

then T_n is the maximum likelihood estimator of θ. Of course, this choice of ψ depends upon knowledge of F_o. In particular, if F_o is the standard normal, then ψ is just $\psi(x) = x$.

Huber (1964) considered modification of this classical formulation in order to obtain *robust* estimation of θ, i.e., in order to obtain a choice of ψ which is optimal when F_o is not completely known. In particular, for the case of F_o assumed to belong to a specified neighborhood of the standard normal, he introduced an associated minimax problem based on the criterion of asymptotic variance of the estimator and obtained as optimal solution a ψ function of the form

$$\psi(x) = \begin{cases} x, & |x| \le k, \\ k\,\mathrm{sgn}(x), & |x| \ge k, \end{cases} \tag{4.4}$$

where k is a parameter determined by the specifications of the minimax problem. Note that this ψ represents a *truncation* of the classical $\psi(x) = x$. In effect, this robust estimator is less sensitive to extreme observations. A number of other ψ functions considered by Huber for the purposes of robust estimation are, like (4.4), *nondecreasing and bounded.*

A further type of modification of the classical $\psi(x) = x$ was introduced

by Hampel (1968), (1974) for the purpose of further reducing the influence of extreme observations on the error of estimation. His ψ functions are not only bounded but in addition "*redescend*" to the origin. A typical version is

$$\psi(x) = -\psi(-x) = \begin{cases} x, & 0 \le x \le a, \\ a, & a \le x \le b, \\ \left(\dfrac{c - x}{c - b}\right)a, & b \le x \le c, \\ 0, & x \ge c. \end{cases} \tag{4.5}$$

Similar in character, but smoother, is the ψ function

$$\psi(x) = -\psi(-x) = \begin{cases} \sin dx, & 0 \le x \le \pi/d \\ 0, & x \ge \pi/d. \end{cases} \tag{4.6}$$

Estimators based on (4.4), (4.5) and (4.6) were studied, along with many other estimators, in the comprehensive Princeton Monte Carlo Study (Andrews *et. al.* (1972)) and were found empirically to be very robust considered with various competitors. Furthermore, "redescenders" have been obtained as optimal solutions in certain robustness problems (see Collins (1976) and Portnoy (1977)).

4.2. *Definition of M-functional.* In formulating a precise notion of the solution of (4.1) as a functional $T(F)$, it is necessary to take into account that the equation (4.1) may have multiple solutions. In the case of a redescending ψ, this complication indeed arises and may entail solutions outside the range of the support of F. It is thus convenient to introduce values $0 < p_1 < p_2 < 1$ such that solutions (4.1) outside the interval $[F^{-1}(p_1), F^{-1}(p_2)]$ may be ignored, where as usual $F^{-1}(p) = \inf\{x: F(x) \ge p\}$. (In practice the judicious selection of p_1 and p_2 is based on the need to satisfy certain assumptions in the lemmas and theorems to follow.)

For a given function $\psi(x)$, $-\infty < x < \infty$, and a given distribution F, we define the associated function

$$\lambda_F(c) = \int_{-\infty}^{\infty} \psi(x - c)\, dF(x), \quad -\infty < c < \infty. \tag{4.7}$$

For given $0 < p_1 < p_2 < 1$, we define the set

$$C(\psi; F; p_1, p_2) = \{c: \lambda_F(c) = 0 \text{ and } F^{-1}(p_1) \le c \le F^{-1}(p_2)\}.$$

DEFINITION. The *M-functional corresponding to* ψ *and* (p_1, p_2) is defined as

$$(4.8) \qquad \begin{aligned} T(F) &= \inf C(\psi; F; p_1, p_2), \text{ if } C(\psi; F; p_1, p_2) \text{ nonempty,} \\ &= F^{-1}(\tfrac{1}{2}(p_1 + p_2)), \text{ otherwise. } \square \end{aligned}$$

Clearly, $T(F)$ takes values only in the interval $[F^{-1}(p_1), F^{-1}(p_2)]$. A more general definition, which we shall not need in the present development, would substitute the condition "$\lambda_F(t)$ changes sign at $t = c$" for the condition "$\lambda_F(c) = 0$." Further approaches toward definition of $T(F)$ are discussed in subsection 4.8 below.

4.3. *Nonstochastic convergence aspects of sequences* $\{T(G_n)\}$. For the case of a sequence of distributions $\{G_n\}$ converging weakly to F, denoted $G \Rightarrow F$, we provide conditions under which $T(G_n) \to T(F)$ as $n \to \infty$. In particular, conditions on ψ are specified.

Define, for $\varepsilon \ge 0$, the open interval

$$(4.9) \qquad I_F(\varepsilon) = (F^{-1}(p_1) - \varepsilon, F^{-1}(p_2 + \varepsilon) + \varepsilon)$$

and denote the closure of $I_F(\varepsilon)$ by $\bar{I}_F(\varepsilon)$. Note that $\bar{I}_F(0)$ contains $T(F)$. The role of $I_F(\varepsilon)$ for $\varepsilon > 0$ is seen in the following result.

LEMMA 4.1. *Suppose that* $G_n \Rightarrow F$. *Let* $0 < p < 1$ *and* $\varepsilon > 0$ *be such that* F *is continuous at* $F^{-1}(p) - \varepsilon$ *and at* $F^{-1}(p + \varepsilon) + \varepsilon$. *(Such choices of* ε, *arbitrarily small, may always be found.) Then, for all sufficiently large* n,

$$(4.10) \qquad F^{-1}(p) - \varepsilon \le G_n^{-1}(p) \le F^{-1}(p + \varepsilon) + \varepsilon.$$

PROOF. Suppose that the inequality $G_n^{-1}(p) < F^{-1}(p) - \varepsilon$ holds for infinitely

many $n = 1, 2, \cdots$. Then the inequalities $p \le G_n(G_n^{-1}(p)) \le G_n(F^{-1}(p) - \varepsilon)$ hold for infinitely many $n = 1, 2, \cdots$. But then the convergence $G_n \Rightarrow F$ yields $p \le F(F^{-1}(p) - \varepsilon)$ and, equivalently, $F^{-1}(p) \le F^{-1}(p) - \varepsilon$, a contradiction. Thus the first inequality in (4.10) holds for all sufficiently large n. The other inequality in (4.10) follows by a similar argument. □

COROLLARY. *Suppose that* $G_n \Rightarrow F$. *Let* $0 < p < 1$. *If* $F^{-1}(\cdot)$ is *continuous at* p, *then* $G_n^{-1}(p) \to F^{-1}(p)$, $n \to \infty$. *That is,* $G_n^{-1} \Rightarrow F^{-1}$.

The preceding lemma shows that if $G_n \Rightarrow F$, then the interval $\bar{I}_{G_n}(0)$, which contains $T(G_n)$, must for large n lie within the interval $\bar{I}_F(\varepsilon)$, which contains T(F). This is one step toward the convergence of $T(G_n)$ to T(F). The next result focuses more directly upon this convergence. To this effect, we introduce a condition which ensures that T(F) is well-defined as a "target" parameter.

CONDITION A. (i) The equation $\lambda_F(c) = 0$ has a *unique* solution T(F) in the interval $[F^{-1}(p_1), F^{-1}(p_2)]$, and $\lambda_F(\cdot)$ changes sign at T(F); (ii) in fact, T(F) lies in the *slightly smaller* interval $(F^{-1}(p_1 + \varepsilon_1), F^{-1}(p_2))$ for some $\varepsilon_1 > 0$; (iii) moreover, T(F) is the unique zero of $\lambda_F(\cdot)$ in the *slightly larger* interval $[F^{-1}(p_1) - \varepsilon_2, F^{-1}(p_2 + \varepsilon_2) + \varepsilon_2]$ for some $\varepsilon_2 > 0$. □

An alternative version of Condition A, which could be substituted for Condition A in the sequel, is given by replacing (ii) by (ii*) $F^{-1}(p)$ is continuous at p_1. Still another version is given by replacing (ii) and (iii), respectively, by (ii') $F^{-1}(p)$ is continuous at p_1 and p_2; (iii') T(F) is the unique zero of $\lambda_F(\cdot)$ in $[F^{-1}(p_1) - \varepsilon, F^{-1}(p_2) + \varepsilon]$ for some $\varepsilon > 0$.

LEMMA 4.2. *Let* F, p_1, p_2 *and* ψ *be such that Condition A holds. Let* $\{G_n\}$ *satisfy*

(4.11) $$G_n \Rightarrow F;$$

(4.12) $\lambda_{G_n}(\cdot)$ *converges continuously to* $\lambda_F(\cdot)$; *i.e., if* $b_n \to b$,

then $\lambda_{G_n}(b_n) \to \lambda_F(b)$, $n \to \infty$;

(4.13) $$\lambda_{G_n}(c) \text{ is continuous in } c, \text{ each } n.$$

Then

(4.14) $$\lim_{n\to\infty} T(G_n) = T(F).$$

PROOF. By A(ii), there exist c_1 and c_2 such that

$$F^{-1}(p_1 + \varepsilon_1) < c_1 < T(F) < c_2 < F^{-1}(p_2).$$

Then, by A(i), $\lambda_F(c_1)$ and $\lambda_F(c_2)$ are opposite in sign. By (4.12) it follows that $\lambda_{G_n}(c_1)$ and $\lambda_{G_n}(c_2)$ have opposite signs for all sufficiently large n. For such n, it follows by (4.13) that the equation $\lambda_{G_n}(c) = 0$ has a solution in the interval (c_1, c_2). By (4.11) and Lemma 4.1, we have, for all sufficiently large n,

$$(c_1, c_2) \subset (G_n^{-1}(p_1), G_n^{-1}(p_2)) \subset (F^{-1}(p_1) - \varepsilon_2, F^{-1}(p_2 + \varepsilon_2) + \varepsilon_2).$$

Thus there exists N_0 such that the set of values $\{T(G_n), n \geq N_0\}$ are solutions of $\lambda_{G_n}(c) = 0$ and lie in the finite interval $I_F(\varepsilon_2)$. Thus there exists a point c_0 in the closed interval $\bar{I}_F(\varepsilon_2)$ such that $T(G_{n_k}) \to c_0$ for some subsequence $\{G_{n_k}\}$, $n_k \geq N_0$. By (4.12) again, we have $\lambda_F(c_0) = 0$. By A(iii), we have $c_0 = T(F)$. We have thus established that every ε-neighborhood of $T(F)$ must contain $T(G_n)$ for all but a finite number of n. That is, (4.14) holds. □

The following result provides various simple conditions on ψ sufficient for one or the other of (4.12) and (4.13) to hold.

LEMMA 4.3. (i) *If* ψ *is continuous and bounded, then* (4.13) *holds. Also,* $\lambda_F(c)$ *is continuous in* c. *If also* (4.11) *holds, then*

(4.15) $$\lambda_{G_n}(b) \to \lambda_F(b),\ n \to \infty, \text{ each } b.$$

(ii) *If* ψ *is continuous and nondecreasing, then* (4.13) *holds. Also,* $\lambda_F(c)$

is continuous in c. *If also* (4.15) *holds, then* (4.12) *holds.* (iii) *If* ψ *is uniformly continuous and* (4.15) *holds, then* (4.12) *holds.*

PROOF. (i) Apply the Dominated Convergence Theorem and the Helly-Bray Theorem. (ii) The first two statements follow from the Monotone Convergence Theorem. Now suppose that (4.15) holds. Let $b_n \to b$. Let $\delta > 0$ be given. Let n be sufficiently large that $|b_m - b| < \delta$, all $m \geq n$. Then, since ψ is non-decreasing, we have

$$(1) \qquad |\lambda_{G_m}(b_n) - \lambda_{G_m}(b)| \leq |\lambda_{G_m}(b + \delta) - \lambda_{G_m}(b)| + |\lambda_{G_m}(b - \delta) - \lambda_{G_m}(b)|.$$

Now

$$|\lambda_{G_m}(b + \delta) - \lambda_{G_m}(b)| \leq |\lambda_F(b + \delta) - \lambda_F(b)| + |\lambda_{G_m}(b + \delta) - \lambda_F(b + \delta)| + |\lambda_{G_m}(b) - \lambda_F(b)|.$$

Thus

$$\sup_{m \geq n} |\lambda_{G_m}(b + \delta) - \lambda_{G_m}(b)| \leq |\lambda_F(b + \delta) - \lambda_F(b)| + B(n, \delta),$$

where

$$B(n, \delta) = \sup_{m \geq n} |\lambda_{G_m}(b + \delta) - \lambda_F(b + \delta)| + \sup_{m \geq n} |\lambda_{G_m}(b) - \lambda_F(b)|.$$

By (4.15), $B(n, \delta) \to 0$ as $n \to \infty$. Hence

$$(2) \qquad \lim_{n\to\infty} \sup_{m \geq n} |\lambda_{G_m}(b + \delta) - \lambda_{G_m}(b)| \leq |\lambda_F(b + \delta) - \lambda_F(b)|.$$

Similarly,

$$(3) \qquad \lim_{n\to\infty} \sup_{m \geq n} |\lambda_{G_m}(b - \delta) - \lambda_{G_m}(b)| \leq |\lambda_F(b - \delta) - \lambda_F(b)|.$$

Combining (1), (2) and (3), we have

$$(4) \qquad \lim_{n\to\infty} \sup_{m \geq n} |\lambda_{G_m}(b_n) - \lambda_{G_m}(b)| \leq |\lambda_F(b + \delta) - \lambda_F(b)| + |\lambda_F(b - \delta) - \lambda_F(b)|.$$

Letting $\delta \to 0$ in the right-hand side of (4), and making use of the continuity of $\lambda_F(c)$ established already, we obtain

(4.16) $$\lim_{n\to\infty} \sup_{m\geq n} |\lambda_{G_m}(b_n) - \lambda_{G_m}(b)| = 0 ,$$

which implies (4.12).

(iii) Write

$$|\lambda_{G_m}(b_n) - \lambda_{G_m}(b)| = |\int[\psi(x - b_n) - \psi(x - b)]dG_m(x)|$$

$$\leq ||\psi(x - b_n) - \psi(x - b)||_\infty .$$

It follows that

(4.17) $$\lim_{n\to\infty} \sup_{m\geq n} |\lambda_{G_m}(b_n) - \lambda_{G_m}(b)| \leq \lim_{n\to\infty} ||\psi(x - b_n) - \psi(x - b)||_\infty = 0 ,$$

which, with (4.15), implies (4.12). □

At this point it is convenient to introduce, for each $\varepsilon > 0$, the set of distributions

$$\mathcal{G}_F(\varepsilon) = \{G: \ \lambda_G(T(G)) = 0 \text{ and } T(G) \in I_F(\varepsilon)\},$$

which will play in the present context of M-functionals the role of the set $\mathcal{G}_F$ in the definition of the quasi-differential. The following result is parallel to Lemma 4.2. Under the same conditions, except that (4.12) is relaxed to (4.15), it provides the slightly weaker conclusion that, for every $\varepsilon > 0$, $T(G_n)$ lies in $I_F(\varepsilon)$ for all sufficiently large n.

LEMMA 4.4. *Let* F, p_1, p_2 *and* ψ *be such that Condition A holds. Let* $\{G_n\}$ *satisfy* (4.11), (4.13) *and* (4.15). *Then, for every* $\varepsilon > 0$,

(4.18) $$G_n \in \mathcal{G}_F(\varepsilon), \text{ all } n \text{ sufficiently large.}$$

PROOF. It is readily seen that if Condition A holds, the constant $\varepsilon_2 > 0$ in A(iii) may be taken arbitrarily small. Now the first part of the proof of Lemma 4.2

establishes that for all sufficiently large n, $\lambda_{G_n}(T(G_n)) = 0$ and $T(G_n) \in I_F(\varepsilon_2)$. Thus we have (4.18). □

4.4. *Strong consistency of* $T(F_n)$. We are now prepared to assert important convergence properties of F_n and $T(F_n)$, analogous to (4.18) and (4.14).

LEMMA 4.5. *Let* F, p_1, p_2 *and* ψ *be such that Condition A holds. Suppose that*

(4.19) $$\psi \text{ is continuous };$$

(4.20) $$||F_n - F||_\infty \xrightarrow{\text{wpl}} 0, \; n \to \infty;$$

(4.21) $$\lambda_{F_n}(c) \xrightarrow{\text{wpl}} \lambda_F(c), \; n \to \infty \text{ (each } c\text{)}.$$

Then, for every $\varepsilon > 0$,

(4.22) $$P\{F_n \in G_F(\varepsilon), \text{ all } n \text{ sufficiently large}\} = 1.$$

If, further, either

(4.23a) $$\psi \text{ is nondecreasing}$$

or

(4.23b) $$\psi \text{ is uniformly continuous},$$

then

(4.24) $$T(F_n) \xrightarrow{\text{wpl}} T(F), \; n \to \infty .$$

PROOF. Writing

(4.25) $$\lambda_{F_n}(c) = \frac{1}{n} \sum_{i=1}^{n} \psi(X_i - c),$$

we see that (4.19) implies (4.13), with G_n replaced by F_n. By (4.20) and (4.21), we have that (4.11) and (4.15), with G_n replaced by F_n, hold with probability 1. Thus, by Lemma 4.4, (4.22) is proved.

If either (4.23a) or (4.23b) holds, then by Lemma 4.3 (ii), (iii) we have

that (4.12), with G_n replaced by F_n, holds with probability 1. Thus (4.24) is proved. □

REMARKS. (i) A sufficient condition for (4.20) and (4.21) is that the X_i's be I.I.D. This follows from the Glivenko-Cantelli Theorem and the classical Strong Law of Large Numbers.

(ii) An alternative sufficient condition for (4.21) is that (4.19) and (4.20) hold and ψ be bounded. This follows by the Helly-Bray Theorem. □

In view of Remark (i) above, we have by Lemma 4.5 the following result characterizing *strong consistency* of $T(F_n)$ under typical conditions on ψ and the X_i's.

THEOREM 4.1. *Let* F, p_1, p_2 *and* ψ *be such that Condition A holds. Suppose that either*

(4.26a) $\quad\psi$ *is continuous and nondecreasing*

or

(4.26b) $\quad\psi$ *is uniformly continuous.*

Let $\{X_i\}$ *be a sequence of independent observations on* F. *Then* (4.22) *and* (4.24) *hold.*

4.5. *A quasi-differential for* $T(\cdot)$. We now investigate the existence of a (possibly stochastic) quasi-differential for an M-functional. Defining

$$H_G(s, t) = \int\psi(x - s)dF_t(x), \tag{4.27}$$

where

$$F_t = F + t(G - F),$$

the equation (4.1) defining $T(F_t)$ may be written

$$H_G(T(F_t), t) = 0. \tag{4.28}$$

By implicit differentiation (with respect to t) in the equation (4.28), we obtain

$$\left.\frac{\partial H}{\partial s}\right|_{s=T(F)} \cdot \left.\frac{dT(F_t)}{dt}\right|_{t=0} + \left.\frac{\partial H}{\partial t}\right|_{t=0} = 0 ,$$

i.e.,

$$\lambda_F'(T(F)) \cdot D_G T(F) + [\lambda_G(T(F)) - \lambda_F(T(F))] = 0 ,$$

i.e., if $\lambda_F(T(F)) = 0$,

$$D_G T(F) = - \frac{\lambda_G(T(F))}{\lambda_F'(T(F))} , \tag{4.29}$$

and, in particular,

$$D_{\delta_x} T(F) = - \frac{\psi(x - T(F))}{\lambda_F'(T(F))} , \ -\infty < x < \infty, \tag{4.30}$$

corresponding to (2.12). We thus define

$$T(F; \Delta) = \frac{\int \psi(x - T(F)) d\Delta(x)}{-\lambda_F'(T(F))} \tag{4.31}$$

and seek to establish that this is a (possibly merely stochastic) quasi-differential for $T(\cdot)$. In particular, we deal with

$$T(F; G - F) = \frac{\int \psi(x - T(F)) d[G(x) - F(x)]}{-\lambda_F'(T(F))} . \tag{4.32}$$

We next express $T(G) - T(F)$ in a form suitable for relating to $T(F; G - F)$. Define

$$h(t) = \frac{\lambda_F(t) - \lambda_F(T(F))}{t - T(F)} , \ t \neq T(F),$$

$$= \lambda_F'(T(F)), \ t = T(F).$$

Then we may write

$$T(G) - T(F) = \frac{\int [\psi(x - T(G)) - \psi(x - T(F))] dF(x)}{h(T(G))} . \tag{4.33}$$

Under the assumption that

$$(4.34) \qquad \lambda_F(T(F)) = 0 \text{ and } \lambda_G(T(G)) = 0,$$

which will be satisfied in the context to be considered, (4.33) takes the form

$$(4.35) \qquad T(G) - T(F) = \frac{\int \psi(x - T(G))d[G(x) - F(x)]}{-h(T(G))} .$$

Comparison of (4.32) and (4.35) reveals compatibility in the numerators but not in the denominators. However, noting that $h(T(G)) \to \lambda_F'(T(F))$ as $T(G) \to T(F)$, we see the utility of the *quasi*-differential approach. Define the functional

$$(4.36) \qquad T_F(G) = \frac{\lambda_F'(T(F))}{h(T(G))} .$$

Then $T_F(F) = 1$ and (4.32) is equivalent to

$$(4.37) \qquad T_F(G)\, T(F; G - F) = \frac{\int \psi(x - T(F))d[G(x) - F(x)]}{-h(T(G))} ,$$

a form somewhat more compatible with (4.35). Indeed, by the use of Lemma 3.1 with (4.35) and (4.37), we arrive at the inequality

$$(4.38) \qquad \frac{|T(G) - T(F) - T_F(G)\, T(F; G - F)|}{||G - F||_\infty} \le \frac{||\psi(x - T(F)) - \psi(x - T(G))||_V}{|h(T(G))|} .$$

(This is still subject to the proviso (4.34).)

The inequality (4.38), in conjunction with the convergence results developed in subsections 4.3 and 4.4, is very useful in characterizing $T(F; \Delta)$ as a quasi-differential or stochastic quasi-differential for $T(\cdot)$. We shall give results covering both versions.

THEOREM 4.2. *Let* F, p_1, p_2 *and* ψ *be such that Condition A holds. Assume that* $\lambda_F'(T(F)) \neq 0$. *Suppose that either*

(4.39a) ψ *is continuous, nondecreasing and bounded*

or

(4.39b) ψ *is uniformly continuous and bounded.*

Suppose also that ψ *satisfies*

(4.40
$$\lim_{b\to 0}||\psi(x-b)-\psi(x)||_V = 0.$$

Then $T(F;\ \Delta)$ *defined by* (4.31) *is a quasi-differential for* $T(\cdot)$ *at* F *with respect to the norm* $||\cdot||_\infty$, *the set* $G_F(\varepsilon_2)$, *and the functional* $T_F(\cdot)$ *defined by* (4.36).

PROOF. Let $\{G_n\}$ satisfy $||G_n - F||_\infty \to 0$, $n\to\infty$. Then (4.11) holds. Hence also, by Lemma 4.3 in conjunction with (4.39), conditions (4.12) and (4.13) hold. Therefore, by Lemmas 4.2 and 4.4, $G_n \in G_F(\varepsilon_2)$ for all n sufficiently large and $T(G_n)\to T(F)$, $n\to\infty$. Thus

(4.41)
$$h(T(G_n)) \to -\lambda_F'(T(F)) \neq 0$$

and (4.34) holds for the given F and for G_n for n sufficiently large. Therefore, by (4.40) and (4.38), we obtain (2.18a). □

We now relax somewhat the requirements on ψ and still obtain that $T(F;\ \Delta)$ is a quasi-differential in the *strong stochastic* sense. A proof similar to that of Theorem 4.2, making use of Theorem 4.1 in place of Lemmas 4.2 and 4.4, yields the following result.

THEOREM 4.3. *Let* F, p_1, p_2 *and* ψ *be such that Condition A holds. Assume that* $\lambda_F'(T(F)) \neq 0$. *Suppose that either*

(4.41a) ψ *is continuous and nondecreasing*

or

(4.41b) ψ *is uniformly continuous.*

Suppose also that ψ *satisfies* (4.40). *Let* $\{X_i\}$ *be a sequence of independent observations on* F. *Then* $T(F;\ \Delta)$ *given by* (4.31) *is a strong stochastic quasi-differential for* $T(\cdot)$ *at* F *with respect to the norm* $\|\cdot\|_\infty$, *the sequence* $\{X_i\}$, *and the functional* $T_F(\cdot)$ *given by* (4.36).

REMARK. Under the conditions of the preceding theorems, the quantity $T(F;\ \Delta)$ given by (4.31) is a (stochastic) quasi-differential *satisfying Condition L.* For later reference, we note

$$T[F;\ X] = -\frac{\psi(x - T(F))}{\lambda_F'(T(F))}, \tag{4.42}$$

$$\mu(T,\ F) = 0, \tag{4.43}$$

and

$$\sigma^2(T,\ F) = \frac{\int_{-\infty}^{\infty} \psi^2(x - T(F))dF(x)}{[\lambda_F'(T(F))]^2}, \tag{4.44}$$

in accord with (2.6), (2.10) and (2.31). □

In subsection 2.4, a concept of differential for a *sequence of statistics,* rather than a *functional,* was introduced. Exploiting this approach, we now further relax the conditions on ψ.

Let T_o be a solution of the equation (4.1), i.e., satisfy $\lambda_F(T_o) = 0$. Analogous to $T(F;\ \Delta)$ given by (4.31), we introduce

$$T_o(F;\ \Delta) = \frac{\int \psi(x - T_o)d\Delta(x)}{-\lambda_F'(T_o)}. \tag{4.45}$$

Analogous to the function $h(t)$ used above, we introduce

$$\begin{aligned} h_o(t) &= \frac{\lambda_F(t) - \lambda_F(T_o)}{t - T_o},\ t \neq T_o, \\ &= \lambda_F'(T_o),\ t = T_o. \end{aligned} \tag{4.46}$$

The role of $T(F_n)$ will now be played by a sequence of statistics $\{T_n\}$ which essentially are solutions of the equation (4.2) and converge to T_o. (Specific

conditions are given in the theorem below.) Corresponding to such a sequence $\{T_n\}$, we define a sequence $\{Z_n\}$ by

$$Z_n = \frac{\lambda_F'(T_o)}{h(T_n)} . \tag{4.47}$$

In the case that T_n is indeed a solution of (4.2), i.e., satisfies

$$\lambda_{F_n}(T_n) = 0, \tag{4.48}$$

we obtain, by a derivation similar to that leading to (4.38),

$$\frac{|T_n - T_o - Z_n T_o(F; F_n - F)|}{||F_n - F||_\infty} \le \frac{||\psi(x - T_o) - \psi(x - T_n)||_V}{|h(T_n)|} . \tag{4.49}$$

The following analogue of Theorem 4.3 thus follows readily.

THEOREM 4.4. *Let* F *and* ψ *be such that* $\lambda_F(t) = 0$ *has a solution* T_o. *Assume that* $\lambda_F'(T_o) \neq 0$. *Suppose that* ψ *is continuous and satisfies* (4.40). *Let* $\{X_i\}$ *be a sequence of independent observations on* F. *Let the sequence* $\{T_n\}$, $T_n = T_n(X_1, \ldots, X_n)$, *satisfy*

$$P\{\lambda_{F_n}(T_n) = 0, \text{ all n sufficiently large}\} = 1 \tag{4.50}$$

and

$$P\{T_n \to T_o, n \to \infty\} = 1. \tag{4.51}$$

Define $\{Z_n\}$ *by* (4.47). *Then* $T_o(F; \Delta)$ *given by* (4.46) *is a strong stochastic quasi-differential for* $\{T_n\}$ *at* (T_o, F) *with respect to the norm* $||\cdot||_\infty$, *the sequence* $\{X_i\}$ *and the sequence* $\{Z_n\}$.

Let us compare the assumptions of Theorems 4.3 and 4.4. In the latter, conditions (4.50) and (4.51) essentially characterize $\{T_n\}$ as, with probability 1, a consistent sequence of solutions of equation (4.2). Implicitly, the same is assumed in Theorem 4.3 for the sequence $\{T(F_n)\}$. This is seen from Theorem 4.1.

Regarding conditions on ψ, Theorem 4.4 retains (4.40) but eliminates (4.41). However, note that (4.40) almost contains the assumption that ψ is continuous. For, noting that

$$(4.52) \qquad |\psi(y_o - b) - \psi(y_o) - [\psi(x_o - b) - \psi(x_o)]| \le ||\psi(x - b) - \psi(x)||_V,$$

and putting $y_o = x_o - b$ in (4.52), we have by (4.40) that

$$\psi(x_o - 2b) - 2\psi(x_o - b) + \psi(x_o) \to 0 \quad \text{as } b \to 0.$$

Thus, if ψ is right- and left-continuous at x_o (as in the case of ψ nondecreasing), then

$$\psi(x_o+) = \psi(x_o-) = \psi(x_o).$$

Moreover, the requirements (4.50) and (4.51) implicitly restrict ψ. Thus Theorem 4.4 represents but a mild relaxation of the assumptions on ψ imposed by Theorem 4.3. On the other hand, Theorem 4.4 provides considerable latitude in the manner of definition of a "solution" of the equation (4.2). Nevertheless Theorem 4.3, while specific to the sequence $\{T(F_n)\}$, does not require a separate treatment of the consistency issue.

4.6. *Asymptotic normality and almost sure behavior of* $T(F_n)$. We have established in Theorem 4.3 that, under appropriate assumptions on ψ and $\{X_i\}$, the M-functional $T(\cdot)$ possesses a strong stochastic quasi-differential. We now apply this in connection with the general results, Theorems 2.1 and 2.2.

Consider the functional $T_F(\cdot)$ defined by (4.36). Under the conditions of Theorem 4.3, it follows by Theorem 4.1 that $T(F_n) \xrightarrow{wpl} T(F)$ and hence, since $h(t)$ is continuous at $t = T(F)$, that

$$(4.53) \qquad T_F(F_n) \xrightarrow{wpl} 1, \; n \to \infty.$$

Therefore, by Theorems 2.1 and 2.2 in conjunction with (4.43), we have

THEOREM 4.5. *Let* F, p_1, p_2 *and* ψ *be such that Condition A holds. Assume that* $\lambda_F'(T(F)) = 0$. *Suppose that either*

(4.54a) ψ *is continuous and nondecreasing*

or

(4.54b) ψ *is uniformly continuous.*

Suppose also that (4.40) *holds. Assume that* $\sigma^2(T, F)$ *given by* (4.44) *is finite and positive. Let* $\{X_i\}$ *be independent observations on* F. *Then*

$$\sqrt{n}\,[T(F_n) - T(F)] \xrightarrow{d} N(0, \sigma^2(T, F)) \tag{4.55}$$

and

$$\overline{\lim_{n\to\infty}} \frac{\sqrt{n}\,[T(F_n) - T(F)]}{\sqrt{2\sigma^2(T, F)\log\log n}} = 1 \text{ w. p. 1.} \tag{4.56}$$

Under the broader assumptions of Theorem 4.4, we have an analogous result, based on

$$\sigma_o^2(T_o, F) = \frac{\int_{-\infty}^{\infty} \psi^2(x - T_o)dF(x)}{[\lambda_F'(T_o)]^2}. \tag{4.57}$$

Specifically, we state

THEOREM 4.6. *Let* F *and* ψ *be such that* $\lambda_F(t) = 0$ *has a solution* T_o. *Assume that* $\sigma_o^2(T_o, F)$ *given by* (4.57) *is finite and positive. Let* $\{X_i\}$ *be independent observations on* F. *Let the sequence* $\{T_n\}$, $T_n = T_n(X_1, \ldots, X_n)$, *satisfy*

$$P\{\lambda_{F_n}(T_n) = 0, \text{ all } n \text{ sufficiently large}\} = 1 \tag{4.58a}$$

and

$$P\{T_n \to T_o,\ n \to \infty\} = 1. \tag{4.58b}$$

Then

$$(4.59) \qquad \sqrt{n}\,[T_n - T_0] \xrightarrow{d} N(0, \sigma_0^2(T_0, F))$$

and

$$(4.60) \qquad \overline{\lim_{n\to\infty}} \frac{\sqrt{n}\,[T_n - T_0]}{\sqrt{2\,\sigma_0^2(T_0, F)\log\log n}} = 1 \quad w.\ p.\ 1.$$

4.7. *Discussion of specific examples.* (i) Regarding restrictions on ψ, note that the classical estimator, $\psi(x) = x$, satisfies (4.54a) and (4.40). So does the "huber" (4.4). The "Hampel" (4.5) and the "sine" (4.6) do not satisfy (4.54a), but they do satisfy (4.54b); they also satisfy (4.40). (In checking the validity of (4.40), a helpful tool is the relation $||g||_V = \int|g'(x)|dx$, valid for absolutely continuous g.) For ψ of the form $\psi(x) = x^\delta$, condition (4.40) holds if and only if $\delta \le 1$.

(ii) Regarding restrictions of F, the major issue is Condition A and its variations. If ψ is continuous and strictly increasing, then (4.1) has at most one solution and virtually any choices of p_1 and p_2 close to 0 and 1, respectively, will allow Condition A to be satisfied. For "redescending" ψ such as (4.5) and (4.6), there are several situations of interest. If the support of F is the real line and F is sufficiently regular (avoiding pathologies and bimodality), then again the solution of (4.1) will be unique. If F has bounded support, multiple solutions may arise, but in typical cases only one solution lies in the range of the support of F. For example, if ψ is given by (4.5) with a = 1.5, b = 3, c = 6, and if F is uniform on (0, 1), then the set of solutions to (4.1) is $\{\frac{1}{2}\} \cup \{x:\ x \le -6\} \cup \{x:\ x \ge 7\}$. In this case Condition A is satisfied for any choices $0 < p_1 < \frac{1}{2} < p_2 < 1$. Continuity of F plays a fairly unimportant role here. Suppose that F is merely in the neighborhood of an "ideal" distribution F_0 such as the standard normal or a uniform. Then for typical ψ the curve $\lambda_F(c)$ follows closely the curve $\lambda_{F_0}(c)$. This follows by Lemma 3.1, which gives

$$(4.61)\qquad |\lambda_F(c) - \lambda_{F_o}(c)| \le ||F - F_o||_\infty \cdot ||\psi||_V.$$

(Note that $||\psi||_V < \infty$ holds for ψ's of the forms (4.4) - (4.6).) Thus, if Condition A is satisfied for F_o, p_1, p_2, then for F in a large portion of a sufficiently small neighborhood of F_o, it holds also for the same p_1, p_2.

4.8. *One-step M-estimators for location.* Discovering solutions to (4.1) or (4.2) is theoretically easy, but the actual computations may be excessively costly. An alternative is to take the first Gauss-Newton iteration of (4.2) as an estimator of T(F) rather than the exact solution $T(F_n)$. Such estimators were included in the Princeton Study (Andrews *et al.* (1972)) and were found to behave very much like their full-iterate counterparts, even in the case of small samples. Bickel (1975) has introduced the use of such estimators in the linear model.

Consider solving $\lambda_F(c) = 0$ using Newton's method with a starting point $\tilde{T}$. The first iteration has the form

$$(4.62)\qquad T^{(1)} = \tilde{T} - \frac{\lambda_F(\tilde{T})}{\lambda_F'(\tilde{T})}.$$

Now suppose that $\lambda_F(T_o) = 0$ and that $\tilde{T}_n$ is an "initial" estimator which converges to T_o in some stochastic sense. We define the *one-step M-estimator* T_n by

$$(4.63)\qquad T_n = \tilde{T}_n - \frac{\lambda_{F_n}(\tilde{T})}{\lambda_{F_n}'(\tilde{T})},$$

where

$$(4.64)\qquad \lambda_{F_n}'(t) = \frac{1}{n}\sum_{i=1}^{n} \psi'(X_i - t).$$

Here we are assuming that ψ' exists everywhere. We could relax this requirement by replacing $\lambda_{F_n}'(\tilde{T}_n)$ in (4.63) by a suitable estimator of $\lambda_F'(T_o)$.

For symmetric distributions F and skew-symmetric ψ, $\psi(x) = -\psi(-x)$, a natural starting estimator would be the sample median $F_n^{-1}(\frac{1}{2})$. In asymmetric situations, there is a problem of finding a starting estimator which actually converges to T_o and satisfies $\lambda_F(T_o) = 0$. Of course, if $\tilde{T}_n \to \tilde{T}_o$ and $\tilde{T}_o$ is close to T_o, then (4.64) would still give a reasonable (though not consistent) estimator of T_o. The following theorem is similar to Theorem 4.4. Let $Z_n = \lambda_F'(T_o)/\lambda_{F_n}'(\tilde{T}_n)$.

THEOREM 4.7. *Let* F, ψ, *and* T_o *be such that* $\lambda_F(T_o) = 0$, $\lambda_F'(T_o) \neq 0$, *and* (4.40) *holds. Suppose that* ψ *and* ψ' *are continuous and that either*

(i) ψ' *satisfies*

(4.65a) $$||\psi'(x - c) - \psi'(x)||_V = O(1),\ c \to 0$$

and

(4.65b) $$\lim_{b \to T_o} \lambda_F'(b) = \lambda_F'(T_o),$$

or

(ii) ψ' *is uniformly continuous.*

Let $\{X_i\}$ *be a sequence of independent observations on* F. *Define* T_n *by* (4.63) *and let* $\tilde{T}_n$ *be an estimator such that*

(4.66) $$\sqrt{n}\,(\tilde{T}_n - T_o) = O_p(1),\ n \to \infty.$$

Then $T(F; \Delta)$ *defined by* (4.45) *is a weak stochastic quasi-differential for* $\{T_n\}$ *at* (T_o, F) *w. r. t.* $||\cdot||_\infty$, $\{X_i\}$, *and* $\{Z_n\}$. *If, further,*

(4.66') $$\tilde{T}_n - T_o = O(||F_n - F||_\infty),\ n \to \infty,\ \textit{w. p. } 1,$$

then $T(F; \Delta)$ *defined by* (4.45) *is a strong stochastic quasi-differential for* $\{T_n\}$ *at* (T_o, F) *w. r. t.* $||\cdot||_\infty$, $\{X_i\}$ *and* $\{Z_n\}$.

In the proof we shall need the following simple extension of the law of large numbers.

LEMMA 4.6. *Let* $\{X_i\}$ *be a sequence of independent observations on* F *and let* $T_n = T_n(X_1, \cdots, X_n)$ *satisfy*

$$T_n \xrightarrow{\text{wp1}} T_o,\ n \to \infty. \tag{4.67}$$

If either

(i) g *is continuous and satisfies*

$$||g(x - c) - g(x)||_V = O(1),\ c \to 0, \tag{4.68a}$$

and

$$\lim_{b \to T_o} \int g(x - b)\, dF(x) = \int g(x - T_o)\, dF(x), \tag{4.68b}$$

or

(ii) g *is uniformly continuous,*

then

$$\frac{1}{n} \sum_{i=1}^{n} g(X_i - T_n) \xrightarrow{\text{wp1}} \int g(x - T_o)\, dF(x),\ n \to \infty. \tag{4.69}$$

PROOF. (i) Write

$$\begin{aligned} \left|\int g(x - T_n)\, dF_n(x) - \int g(x - T_o)\, dF(x)\right| \le & \left|\int [g(x - T_n) - g(x - T_o)]\, d[F_n(x) - F(x)]\right| \\ & + \left|\int [g(x - T_n) - g(x - T_o)]\, dF(x)\right| \\ & + \left|\int g(x - T_o)\, d[F_n(x) - F(x)]\right|. \end{aligned}$$

By Lemma 3.1, the first term is bounded by $||F_n - F||_\infty ||g(x - T_n) - g(x - T_o)||_V$, which $\xrightarrow{\text{wp1}} 0$ by (4.68a) and the Glivenko-Cantelli Theorem. The second term $\xrightarrow{\text{wp1}} 0$ by (4.68b), and the third term $\xrightarrow{\text{wp1}} 0$ by the classical SLLN.

(ii) Write

$$\left|\int g(x - T_n)\, dF_n(x) - \int g(x - T_o)\, dF(x)\right| \le \left|\int [g(x - T_n) - g(x - T_o)]\, dF_n(x)\right|$$

$$+ \left|\int g(x - T_o)\, d[F_n(x) - F(x)]\right|.$$

The first term is bounded by $\|g(x - T_n) - g(x - T_o)\|_\infty$, which $\xrightarrow{\text{wpl}} 0$ by the uniform continuity of g. The second term $\xrightarrow{\text{wpl}} 0$ by the SLLN. □

REMARK. Condition (4.68a) is a relaxation of (4.40) (for $g = \psi$) from $o(1)$ to $O(1)$. If (4.67) is replaced by $T_n \xrightarrow{p} T_o$, then the analogous weak law follows. □

PROOF OF THEOREM 4.7. We will prove the "strong" version, the "weak" version following analogously. By appropriate substitutions for T_n, Z_n and $T(F; F_n - F)$, we have

$$|T_n - T_o - Z_n T(F; F_n - F)| = \left|\tilde{T}_n - T_o - \frac{\lambda_{F_n}(\tilde{T}_n)}{\lambda'_{F_n}(\tilde{T}_n)} - \frac{\lambda_{F_n}(T_o) - \lambda_F(T_o)}{-\lambda'_{F_n}(\tilde{T}_n)}\right|$$

(4.70)

$$\le \left|\tilde{T}_n - T_o - \frac{\lambda_F(\tilde{T}_n)}{\lambda'_{F_n}(\tilde{T}_n)}\right| + \left|\frac{1}{\lambda'_{F_n}(\tilde{T})}\right| \cdot \left|\int [\psi(x - \tilde{T}_n) - \psi(x - T_o)]\, d[F_n(x) - F(x)]\right|.$$

The use of either (i) or (ii), in conjunction with (4.66') and Lemma 4.6 with $g = \psi'$, yields

$$\lambda'_{F_n}(\tilde{T}_n) \xrightarrow{\text{wpl}} \lambda'_F(T_o),\ n \to \infty.$$

Since $\lambda'_F(T_o) \ne 0$, the second term of (4.70) is $o(\|F_n - F\|_\infty)$, by (4.40). Now take h_o as in (4.46), so that

(4.71) $$h_o(\tilde{T}_n)(\tilde{T}_n - T_o) = \lambda_F(\tilde{T}_n) - \lambda_F(T_o) = \lambda_F(\tilde{T}_n).$$

Substitution for $\lambda_F(\tilde{T}_n)$ in the first term of (4.70) gives

$$\left|\tilde{T}_n - T_o - \frac{\lambda_F(\tilde{T}_n)}{\lambda'_{F_n}(\tilde{T}_n)}\right| = \left|\tilde{T}_n - T_o - \frac{h_o(\tilde{T}_n)(T_n - T_o)}{\lambda'_{F_n}(\tilde{T}_n)}\right|$$

$$= |\tilde{T}_n - T_o| \cdot \left|1 - \frac{h_o(\tilde{T}_n)}{\lambda'_{F_n}(\tilde{T}_n)}\right|.$$

This last expression is easily seen to be $o(||F_n - F||_\infty)$ w. p. 1 since $\tilde{T}_n - T_o = O(||F_n - F||_\infty)$ and $h_o(\tilde{T}_n)/\lambda'_{F_n}(\tilde{T}_n) \xrightarrow{wp1} 1$, $n \to \infty$. □

The following theorem uses the results of Theorem 4.7 to provide asymptotic normality and the law of the iterated logarithm for the one-step M-estimator T_n defined by (4.64).

THEOREM 4.8. *Let* F, ψ, *and* T_o *be such that* $\lambda_F(T_o) = 0$, $\lambda'_F(T_o) \neq 0$, *and* (4.40) *holds. Suppose that* ψ *and* ψ' *are continuous and also that* ψ' *satisfies* (4.65) *or is uniformly continuous. Suppose that* $\sigma_o^2(T_o, F)$ *given by* (4.58) *is finite and positive. Let* $\{X_i\}$ *be a sequence of independent observations on* F. *Define* T_n *by* (4.63) *and let* $\tilde{T}_n$ *satisfy* (4.66). *Then*

$$\sqrt{n}\,(T_n - T_o) \xrightarrow{d} N(0, \sigma_o^2(T_o, F)),\ n \to \infty. \tag{4.72}$$

If, further, (4.66') *holds, then*

$$\overline{\lim_{n\to\infty}} \frac{\sqrt{n}\,(T_n - T_o)}{\sqrt{2\sigma_o^2(T_o, F)\log\log n}} = 1 \quad w.\ p.\ 1. \tag{4.73}$$

EXAMPLE. Consider ψ_o found by Collins (1976), Theorem 3.1, as the solution of a minimax problem:

$$\psi_o(x) = -\psi_o(-x) = \begin{cases} x,\ 0 \le x \le x_o, \\ x_1 \tanh[\tfrac{1}{2}x_1 (c - x)],\ x_o \le x \le c, \\ 0,\ x > c, \end{cases}$$

where x_o and x_1 are related by $x_o = x_1 \tanh[\frac{1}{2}x_1(c - x_o)]$. Suppose that F is symmetric about θ (playing the role of T_o). Then $\lambda_F(\theta) = 0$, $\lambda_F'(\theta) \neq 0$, (4.40) holds, and ψ' is uniformly continuous. Let $\tilde{T}_n = F_n^{-1}(\frac{1}{2})$. Then, as is well-known, (4.66) holds. If, further, $F'(\theta) > 0$, then (4.66') also follows. Smoothed versions of (4.4) - (4.6) provide other possibilities for ψ. □

4.9 *M-estimators, scale unknown.* In previous subsections the motivating statistical setting for M-estimation was the simple location problem. Thus the functionals defined were invariant w. r. t. changes in location (i.e., $T(F(x - \mu) = T(F) + \mu)$, but not necessarily w. r. t. changes in scale. For practical use we almost always need location functionals which are also scale invariant (i.e., $T(F((x - \mu)/a)) = aT(F) + \mu$, $a > 0$). The obvious way to achieve the desired invariance for M-functionals is to define $T(F)$ as a solution of

$$(4.74) \qquad \int \psi\left(\frac{x - T}{S(F)}\right) dF(x) = 0,$$

where $S(F)$ is a scale functional satisfying $S(F((x - \mu)/a)) = aS(F)$ for $a > 0$ and all μ. Then if $T(F)$ is a solution of (4.74), we have

$$(4.75) \qquad \int \psi\left(\frac{x - (aT(F) + \mu)}{S(F(\frac{x - \mu}{a}))}\right) dF\left(\frac{x - \mu}{a}\right) = \int \psi\left(\frac{y - T(F)}{S(F)}\right) dF(x) = 0,$$

showing that $T(F((x - \mu)/a)) = aT(F) + \mu$. One may choose $S(F)$ independently or obtain $S(F)$ by simultaneously solving

$$\int \psi\left(\frac{x - T}{S}\right) dF(x) = 0$$

and

$$\int \chi\left(\frac{x - T}{S}\right) dF(x) = 0,$$

where, e.g., $\chi(t) = t\psi(t) - 1$ (see Bickel (1975) or Huber (1964)). We prefer

to choose S(F) independently, e.g., the interquartile range

$$S(F) = \frac{F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4})}{2},$$

or any member of a large class of linear functions of order statistics for scale estimation. In either case we would expect $S(F_n) \to S(F)$ to hold in some stochastic sense. The natural estimator $T(F_n)$ would then be a solution of

$$(4.76) \qquad \int \psi\left(\frac{x - T}{S(F_n)}\right) dF_n(x) = 0.$$

First we make precise the definition of M-functional $T(\cdot)$ for the scale unknown case. Then we give analogues of Lemmas 4.2, 4.3 and 4.4 and Theorems 4.1 and 4.4.

As previously, we allow for possible nonuniqueness of solution in (4.74) and (4.75), by selecting the smallest solution lying in $\overline{I}_F(0) = [F^{-1}(p_1), F^{-1}(p_2)]$. We introduce the following definitions:

$$(4.77) \qquad \lambda_F(c, s) = \int_{-\infty}^{\infty} \psi\left(\frac{x - c}{s}\right) dF(x).$$

$$(4.78) \quad C_s(\psi; F; p_1; p_2) = \{c: \lambda_F(c, S(F)) = 0 \text{ and } F^{-1}(p_1) \le c \le F^{-1}(p_2)\}.$$

$$(4.79) \qquad T(F) = \inf C_s(\psi; F; p_1; p_2), \text{ if } C_s(\psi; F; p_1, p_2) \text{ nonempty},$$
$$= F^{-1}(\tfrac{1}{2}(p_1 + p_2)), \text{ otherwise}.$$

$$(4.80) \qquad G_F^s(\varepsilon) = \{G: \lambda_G(T(G), S(G)) = 0 \text{ and } T(G) \in I_F(\varepsilon)\}.$$

CONDITION A_s. (i) The equation $\lambda_F(c, S(F)) = 0$ has a unique solution $c = T(F)$ in the interval $[F^{-1}(p_1), F^{-1}(p_2)]$, and $\lambda_F(c, S(F))$ changes sign at $T(F)$;

(ii) In fact, T(F) lies in the *slightly smaller* interval $(F^{-1}(p_1 + \varepsilon_1), F^{-1}(p_2))$ for some $\varepsilon_1 > 0$;

(iii) Moreover, $T(F)$ is the unique zero of $\lambda_F(\cdot, S(F))$ in the *slightly larger* interval $[F^{-1}(p_1) - \varepsilon_2, F^{-1}(p_2 + \varepsilon_2) + \varepsilon_2]$ for some $\varepsilon_2 > 0$. □

The following analogue of Lemma 4.2 follows by similar arguments.

LEMMA 4.7. *Let* $S(\cdot)$, F, p_1, p_2, *and* ψ *be such that Condition* A_s *holds. Let* $\{G_n\}$ *satisfy* (4.11) *and*

(4.81) $$S(G_n) \to S(F),\ n \to \infty;$$

(4.82) $\lambda_{G_n}(\cdot,\cdot)$ *converges continuously to* $\lambda_F(\cdot,\cdot)$; *i.e., if* $(c_n, d_n) \to (c, d)$, *then* $\lambda_{G_n}(c_n, d_n) \to \lambda_F(c, d)$, $n \to \infty$;

(4.83) $\lambda_{G_n}(c, s)$ *is continuous in* c, *each* s *and each* n.

Then

(4.84) $$\lim_{n\to\infty} T(G_n) = T(F).$$

A stronger condition than (4.83) is

(4.85) $\lambda_{G_n}(c, s)$, *each* n, and $\lambda_F(c, s)$ are each *jointly continuous* in (c, s).

The next result, analogous in part to Lemma 4.3, gives sufficient conditions for (4.85) and (4.82) to hold.

LEMMA 4.8. (i) *If* ψ *is continuous and bounded, then* (4.85) *holds. If also* (4.11) *holds, then*

(4.86) $$\lambda_{G_n}(c, s) \to \lambda_F(c, s) \text{ as } n \to \infty, \text{ all } (c, s).$$

(ii) *If* ψ *is continuous and nondecreasing, then* (4.85) *holds. If also* (4.11), (4.81), *and* (4.86) *hold, then* (4.82) *holds.*

(iii) ***If*** ψ ***is continuous and of bounded variation, then*** (4.85) *holds. If also* $||G_n - F||_\infty \to 0$ *as* $n \to \infty$, *then* (4.86) *and* (4.82) *hold.*

REMARK. Since (iii) of Lemma 4.3 assumes ψ to be uniformly continuous, one might expect (iii) of the above analogous lemma to make the same assumption. However, in proving (iii) of Lemma 4.3, we take advantage of the fact that $|(x - c_n) - (x - c)| = |c_n - c|$ is independent of x, whereas for the above situation, $\left|\psi\left(\frac{x - c_n}{d_n}\right) - \psi\left(\frac{x - c}{d}\right)\right|$ depends on x. Thus the stronger assumption of bounded variation is made. □

PROOF. (i) Apply the Dominated Convergence Theorem and the Helly-Bray Theorem. (ii) The first statement follows from the Monotone Convergence Theorem. Now suppose (4.11), (4.81) and (4.86) hold. Let $\delta_1 > 0$ and $\delta_2 > 0$ be given. Choose n_o large enough that $|c_n - c| < \delta_1$ and $|S(G_n) - S(F)| < \delta_2$ for all $n \geq n_o$. Then, since ψ is nondecreasing, we have

$$|\lambda_{G_n}(c_n, S(G_n)) - \lambda_{G_n}(c, S(F))| \leq |\lambda_{G_n}(c + \delta_1, S(F) + \delta_2) - \lambda_{G_n}(c, S(F))|$$

$$+ |\lambda_{G_n}(c - \delta_1, S(F) - \delta_2) - \lambda_{G_n}(c, S(F))|.$$

Continuing as in the proof of Lemma 4.3, we have

$$\lim_{n\to\infty} |\lambda_{G_n}(c_n, S(G_n)) - \lambda_{G_n}(c, S(F))| = 0. \tag{4.87}$$

Then (4.82) follows from (4.86) and (4.87) since

$$|\lambda_{G_n}(c_n, S(G_n)) - \lambda_F(c, S(F))| \leq |\lambda_{G_n}(c_n, S(G_n)) - \lambda_{G_n}(c, S(F))|$$

$$+ |\lambda_{G_n}(c, S(F)) - \lambda_F(c, S(F))|.$$

(iii) Since ψ is bounded, (4.85) follows from (i). Since the condition $||G_n - F||_\infty \to 0$ implies $G_n \Rightarrow F$, (4.86) also follows from (i). Then

$$|\lambda_{G_n}(c_n, S(G_n)) - \lambda_F(c, S(F))| \le |\lambda_{G_n}(c_n, S(G_n)) - \lambda_F(c_n, S(G_n))|$$

$$+ |\lambda_F(c_n, S(G_n)) - \lambda_F(c, S(F))|.$$

Lemma 3.1 allows us to bound the first term by $||G_n - F||_\infty \cdot ||\psi||_V$. The second term converges to 0 as $n \to \infty$ by (4.85). □

LEMMA 4.9. *Let* $S(\cdot)$, F, p_1, p_2 *and* ψ *be such that Condition* A_s *holds. Suppose that* $\{X_i\}$ *is a sequence of observations on* F *such that*

(4.88) $$S(F_n) \xrightarrow{\text{wpl}} S(F),\ n \to \infty;$$

(4.89) $$||F_n - F||_\infty \xrightarrow{\text{wpl}} 0,\ n \to \infty;$$

(4.90) $$\lambda_{F_n}(c, s) \xrightarrow{\text{wpl}} \lambda_F(c, s),\ n \to \infty\ (\textit{each } (c, s)).$$

Suppose that either

(4.91) ψ *is continuous and nondecreasing*

or

(4.92) ψ *is continuous and of bounded variation.*

Then

(4.93) $$P\{F_n \in G_F^s(\epsilon_2),\ \textit{all } n \textit{ sufficiently large}\} = 1$$

and

(4.94) $$T(F_n) \xrightarrow{\text{wpl}} T(F),\ n \to \infty.$$

PROOF. Writing

(4.95) $$\lambda_{F_n}(c, s) = \frac{1}{n}\sum_{i=1}^{n} \psi\left(\frac{X_i - c}{s}\right),$$

we see that ψ continuous implies (4.83), with G_n replaced by F_n. By (4.89) and (4.90), we have that (4.11), (4.81), and (4.86), with G_n replaced by F_n, hold

w. p. 1. Then by (4.91) or (4.92) (via Lemma 4.8), we have that (4.82), with G_n replaced by F_n, holds w. p. 1. Thus by Lemma 4.7, (4.93) and (4.94) hold. □

THEOREM 4.9. *Let* $S(\cdot)$, F, p_1, p_2, *and* ψ *be such that Condition* A_s *holds. Suppose that either* (4.91) *or* (4.92) *holds. Let* $\{X_i\}$ *be a sequence of independent observations on* F *such that*

$$S(F_n) \xrightarrow{\text{wpl}} S(F), \ n \to \infty.$$

Then (4.93) *and* (4.94) *hold.*

PROOF. Conditions (4.89) and (4.90) hold by the Glivenko-Cantelli Theorem and the classical Strong Law of Large Numbers. □

We now establish a differential in the scale unknown case. Omitting analogues to Theorems 4.2 and 4.3, the most general result is provided, an analogue to Theorem 4.4. Some further definitions are necessary. Denote the partial derivatives of $\lambda_F(c, s)$ at (T_o, S_o) by $D_1\lambda_F(T_o, S_o)$ and $D_2\lambda_F(T_o, S_o)$. Define h_o^* by

$$h_o^*(t) = \frac{\lambda_F(t, S_o) - \lambda_F(T_o, S_o)}{t - T_o}, \ t \neq T_o$$

$$= D_1\lambda_F(T_o, S_o), \ t = T_o.$$

In the following theorem, T_n is arbitrary; one possibility is $T_n = T(F_n)$, for $T(\cdot)$ defined by (4.79). Let $Z_n = D_1\lambda_F(T_o, S_o)/h_o^*(T_n)$.

THEOREM 4.10. *Let* F *and* ψ *be such that* $\lambda_F(t, S_o)$ *has a solution* T_o. *Assume that* $D_1\lambda_F(T_o, S_o) \neq 0$ *and that* $D_2\lambda_F(t, s)$ *is continuous in a neighborhood of* (T_o, S_o). *Suppose that* ψ *is continuous,* that (4.40) *holds,* and that

$$\lim_{c \to 1} ||\psi(cx) - \psi(x)||_V = 0. \tag{4.96}$$

Let $\{X_i\}$ *be a sequence of independent observations on* F. *Let* $T_n = T_n(X_1, \ldots, X_n)$

and $S_n = S_n(X_1, \ldots, X_n)$ *be estimators such that*

(4.97) $\{S_n\}$ *has a strong stochastic differential* $S(F; \Delta)$ *at* (S_o, F) *w. r. t.* $||\cdot||_\infty$ *and the sequence* $\{X_i\}$;

(4.98) $$S_n - S_o = O(||F_n - F||_\infty),\ n \to \infty,\ \text{w. p. 1};$$

(4.99) $$T_n \xrightarrow{\text{wpl}} T_o,\ n \to \infty;$$

(4.100) $$P\{\lambda_{F_n}(T_n, S_n) = 0,\ \textit{all}\ n\ \textit{sufficiently large}\} = 1.$$

Then $\{T_n\}$ *has a strong stochastic quasi-differential* $T(F; \Delta)$ *at* (T_o, F) *w. r. t.* $||\cdot||_\infty$, *the sequence* $\{X_i\}$, *and the sequence* $\{Z_n\}$, *given by*

(4.101) $$T(F; \Delta) = \frac{\int \psi\left(\frac{x - T_o}{S_o}\right) d\Delta(x) + S(F; \Delta)\, D_2\lambda_F(T_o, S_o)}{-D_1\lambda_F(T_o, S_o)}$$

PROOF. For n sufficiently large, $\lambda_F(T_o, S_o) = \lambda_{F_n}(T_n, S_n) = 0$ w. p. 1, and thus w. p. 1,

$$\begin{aligned}
h_o^*(T_n)(T_n - T_o) &= \lambda_F(T_n, S_o) - \lambda_F(T_o, S_o) \\
&= \lambda_F(T_n, S_o) - \lambda_{F_n}(T_n, S_n) \\
&= \lambda_F(T_n, S_o) - \lambda_{F_n}(T_n, S_o) + \lambda_{F_n}(T_n, S_o) - \lambda_{F_n}(T_n, S_n) \\
&= \int \psi\left(\frac{x - T_n}{S_o}\right) d(F(x) - F_n(x)) + \lambda_F(T_n, S_o) - \lambda_F(T_n, S_n) \\
&\quad + \int \left\{\psi\left(\frac{x - T_n}{S_o}\right) - \psi\left(\frac{x - T_n}{S_n}\right)\right\} d(F_n(x) - F(x)).
\end{aligned}$$

Also,

$$h_o^*(T_n)\, Z_n T(F;\, F_n - F) = -\left(\int \psi\left(\frac{x - T_o}{S_o}\right) d(F_n(x) - F(x)) + S(F;\, F_n - F)\, D_2\lambda_F(T_o,\, S_o)\right).$$

Then, for n sufficiently large, w. p. 1,

$$
\begin{aligned}
|h_o^*(T_n)|\, |T_n - T_o - Z_n T(F;\, F_n - F)| \le\ & \left|\int \left[\psi\left(\frac{x - T_n}{S_o}\right) - \psi\left(\frac{x - T_o}{S_o}\right)\right] d(F_n(x) - F(x))\right| \\
& + \left|\int \left[\psi\left(\frac{x - T_n}{S_o}\right) - \psi\left(\frac{x - T_n}{S_n}\right)\right] d(F_n(x) - F(x))\right| \qquad (4.102) \\
& + |\lambda_F(T_n,\, S_o) - \lambda_F(T_n,\, S_n) + S(F;\, F_n - F) D_2\lambda_F(T_o - S_o)|.
\end{aligned}
$$

The first two terms of (4.102) are $o(||F_n - F||_\infty)$ as $n \to \infty$, w. p. 1, by (4.40) and (4.96). By the mean value theorem, for n sufficiently large,

$$\lambda_F(T_n,\, S_n) - \lambda_F(T_n,\, S_o) = D_2\lambda_F(T_n^*,\, S_n^*)(S_n - S_o) \text{ w. p. 1,}$$

where T_n^* and S_n^* are such that $T_n^* \xrightarrow{wpl} T_o$ and $S_n^* \xrightarrow{wpl} S_o$. Thus, the third term of (4.102) can be bounded by

$$|D_2\lambda_F(T_n^*,\, S_n^*) - D_2\lambda_F(T_o,\, S_o)|\, |S_n - S_o| + |S_n - S_o - S(F;\, F_n - F)|.$$

Condition (4.97) and (4.98) then yield the desired $o(||F_n - F||_\infty)$ w. p. 1 result. Division by $h_o^*(T_n)$ poses no problem since $h_o^*(T_n) \xrightarrow{wpl} D_1\lambda_F(T_o,\, S_o) \neq 0$. □

COROLLARY. *Let the hypotheses of Theorem 4.10 hold. Suppose that* σ_o^2 *is finite and positive, where* σ^2 *is given by*

$$\sigma^2 = \mathrm{Var}_F\{\psi\left(\frac{X - T_o}{S_o}\right) + S(F;\, S_X - F) D_2\lambda_F(T_o,\, S_o)\}(D_1\lambda_F(T_o,\, S_o))^{-2}.$$

Then

$$\sqrt{n}(T_n - T_o) \xrightarrow{d} N(0,\, \sigma_o^2) \text{ as } n \to \infty$$

and

$$\overline{\lim_{n\to\infty}} \frac{\sqrt{n}\,(T_n - T_o)}{\sqrt{2\,\sigma_o^2 \log\log n}} = 1 \quad w.\ p.\ 1.$$

REMARKS. (i) If $D_2\lambda_F(T_o, S_o) = 0$, then (4.97) is unnecessary.

(ii) We could weaken the convergences in (4.97)-(4.100) and conclude that (4.101) is a weak stochastic quasi-differential for $\{T_n\}$.

(iii) the ψ functions given by (4.4) and (4.5) satisfy (4.96). □

4.10. *Comparisons with other results.*

The introduction of M-estimation as a formal approach is due to Huber (1964), who defines T_n to be any representative taken from the set of solutions of the equation (4.2). For the case that ψ is nondecreasing, he establishes strong consistency and asymptotic normality of T_n. A parallel result is given by our Theorem 4.5 with the option (4.54a). Our conditions on ψ are slightly more stringent, but we provide in addition the law of the iterated logarithm (LIL). Huber gives another theorem which establishes asymptotic normality of T_n in the case that ψ has a uniformly continuous derivative and under the assumption that consistency of T_n has already been shown by some method. Our Theorem 4.6 is comparable to this result. Not only is our condition on ψ milder in nature, but also we obtain the LIL in addition to the asymptotic normality.

Of course, the main objective of our investigation into M-functionals has been to obtain useful new results not for "Hubers" but for the "redescenders" introduced by Hampel (1968), (1974). These statistics are not adequately handled by Huber's treatment. Nor does Hampel's treatment take up questions of asymptotic normality and almost sure behavior. However, Collins (1976) establishes asymptotic normality of T_n for the case that ψ is continuous with continuous derivative ψ' and skew-symmetric and vanishes outside an interval [-c, c], and F

is governed by the standard normal density on an interval $[T_o - d, T_o + d]$, $d > c$. (Outside this interval, F is allowed to be arbitrary.) Collins takes T_n to be the solution of (4.2) obtained by Newton's method starting with the sample median. The almost sure behavior of T_n is not treated. Our Theorem 4.5 with the option (4.54b) provides a parallel result. The restrictions on ψ are of roughly the same strength, although we do not require ψ' to exist everywhere, but we greatly relax the requirements on F and we characterize the almost sure behavior of $T(F_n)$. Collins also extends his asymptotic normality result to the scale unknown case. The corollary to Theorem 4.10 provides a parallel result, though here we essentially require ψ' to exist everywhere. Again, we also provide the almost sure behavior of $T(F_n)$.

An investigation of Carroll (1975), (1977) treats Behadur-type (see Bahadur (1966)) almost sure asymptotic representations for M-estimators. He requires that ψ be bounded and uniformly Lipschitz of order 1 and possess two continuous bounded derivatives piecewise on intervals, and that F be Lipschitz in neighborhoods of the endpoints of these intervals. He further requires that T_n be strongly consistent for T_o (which entails implicitly further conditions on ψ and F). The desired representation, once established, is quite fruitful, yielding asymptotic normality and the LIL as by-products.

Bickel (1975) treats one-step M-estimators in the general linear model. One of his alternative conditions (E_1, p. 432) requires ψ' to be uniformly continuous. He provides asymptotic normality but does not consider the almost sure behavior of $T(F_n)$.

Portnoy (1977) establishes asymptotic normality of T_n in the case that ψ is bounded and has a derivative which is bounded and uniformly continuous except on a Lebesque-null set and that F is continuous and symmetric and has a density satisfying certain regularity properties. The estimator T_n is taken to be the

solution of (4.2) nearest to a given consistent estimator $\tilde{\theta}_n$. The almost sure behavior of T_n is not treated. However, Portnoy does allow m-dependence rather than strict independence.

Beran (1977a) establishes Hellinger metric differentiability of M-functionals in the case that ψ is strictly monotone and bounded, $\lim_{x\to\infty} \psi(x) > 0$, $\lim_{x\to-\infty} \psi(x) < 0$, and ψ has a continuous bounded derivative.

REFERENCES

Andrews, D. F. *et al.* (1972). *Robust Estimates of Location.* Princeton Univ. Press.

Apostol, T. M. (1957). *Mathematical Analysis.* Addison-Wesley, Reading, Mass.

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* 37 577-580.

Beran, R. (1977a). Robust location estimates. *Ann. Statist.* 5 431-444.

Beran, R. (1977b). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* 5 445-463.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* 70 428-434.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Boos, D. D. (1977a). A differential for L-statistics. FSU Statistics Report M416 (ONR Technical Report No. 116), Florida State University, Tallahassee.

Boos, D. D. (1977b). The differential approach in statistical theory and robust inference. Ph.D. dissertation, Florida State University, Tallahassee.

Carroll, R. J. (1975). An asymptotic representation for M-estimators and linear functions of order statistics. Institute of Statistics Mimeo Series #1006, University of North Carolina, Chapel Hill.

Carroll, R. J. (1977). An almost sure invariance principle for M-estimates. Unpublished manuscript.

Chung, K. L. (1968). *A Course in Probability Theory.* Second Edition. Academic Press, New York.

Collins, J. R. (1976). Robust estimation of a location parameter in the presence of asymmetry. *Ann. Statist.* 4 68-85.

Dieudonne, J. (1960). *Foundations of Modern Analysis.* Academic Press, New York.

Dunford, N. and Schwartz, J. T. (1958). *Linear Operators.* Vol. I. Interscience Publishers, New York.

Filippova, A. A. (1962). Mises theorem on the asymptotic behavior of functionals of empirical distribution functions and statistical applications. *Theor. Probability Appl.* 7 24-57.

Gregory, G. G. (1976). Large sample theory of differentiable statistical functions. Unpublished manuscript.

Hampel, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Univ. of California, Berkeley.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* 69 383-393.

Hartman, P. and Wintner, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.* 63 169-176.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19 293-325.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35 73-101.

Kallianpur, G. and Rao, C. R. (1955). On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhya* 15 331-342.

Kiefer, J. and Wolfowitz, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* 87 173-186.

Kolmogorov, A. N. (1933). Sulla determinzione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Altuari* 4 83-91.

Natanson, I. P. (1961). *Theory of Functions of a Real Variable.* Revised Edition. Ungar, New York.

Portnoy, S. L. (1977). Robust estimation in dependent situations. *Ann. Statist.* 5 22-43.

Richter, H. (1974). Das Gesetz vom iterierten Logarithmus für empirische Verteilungsfunktionen im R^k. *Manuscripta Math.* 11 291-303.

Schmid, P. (1958). On the Kolmogorov and Smirnov limit theorems from discontinuous distribution functions. *Ann. Math. Statist.* 29 1011-1027.

Serfling, R. J. (1971). The law of the iterated logarithm for U-statistics and related von Mises statistics. (abstract) *Ann. Math. Statist.* 42 1794.

von Mises, R. (1947). On the asymptotic distributions of differentiable statistical functions. *Ann. Math. Statist.* 18 309-348.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M415 ONR No. 115	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE On Differentials, Asymptotic Normality and Almost Sure Behavior of Statistical Functions, with Application to M-statistics for Location Parameters		5. TYPE OF REPORT & PERIOD COVERED Technical Report PERFORMING ORG. REPORT NUMBER FSU Statistics Report M415
7. AUTHOR(s) Dennis D. Boos R. J. Serfling		8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-76-C-0608
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics & Probability Program Arlington, Virginia 22217		12. REPORT DATE May, 1977 13. NUMBER OF PAGES 64
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

19. KEY WORDS

Differentials, Statistical Functions; Asymptotic Normality; Almost Sure Behavior; M-Estimates.

20. ABSTRACT

Parameters of interest in statistics can often be expressed as functionals $T(F)$ of the underlying population distribution function, in which case a natural sample analogue estimator is provided by the "statistical function" $T(F_n)$ based upon the sample distribution function F_n.

F sub n

Several notions of differentiability of functionals T are formulated, including innovations designed to broaden the scope of statistical application. Methodology for finding the differential, and for utilizing it to characterize the asymptotic distribution and almost sure behavior of statistical functions, is presented. Typically this means asymptotic normality and the law of the iterated logarithm. Previous work of von Mises (1947), Kallianpur and Rao (1955), Filippova (1962), Gregory (1976) and Beran (1977) is relevant.

Application to M-estimates for location parameters is carried out. The solution of the equation $\int\psi(x-T(F))dF(x) = 0$ is formulated as an M-functional and conditions for its differentiability are investigated. Asymptotic normality and the law of the iterated logarithm for $T(F_n)$ are established under regularity conditions on ψ slightly stronger than continuity and under minimal restrictions on F. One-step estimators and the case of scale unknown are also treated. Previous work of Huber (1964), Hampel (1974), Carroll (1975, 1977), Collins (1976), Portnoy (1977) and Beran (1977) is augmented.